Non-Enzymatic Copying of Nucleic Acid Templates

Non-Enzymatic Copying of Nucleic Acid Templates

A dissertation presented

by

Jonathan Craig Blain

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University

Cambridge, Massachusetts

September 2013

Dissertation Advisor: Professor Jack W. Szostak Jonathan Craig Blain

Non-Enzymatic Copying of Nucleic Acid Templates

Abstract

All known living cells contain a complex set of molecular machinery to support their growth and replication. However, the earliest cells must have been much simpler, consisting of a compartment and a genetic material to allow for Darwinian evolution. To study these intermediates, plausible model 'protocells' must be synthesized in the laboratory since no fossils remain. Recent work has shown that fatty acids can self-assemble into vesicles that are able to grow and divide through simple mechanisms. However, a self-replicating protocell genome has not yet been developed. Here we discuss studies of systems that allow for the copying of nucleic acid templates without enzymes and how they could be developed into a genetic material.

The simplest method of non-enzymatic template copying is to use activated nucleotides that can polymerize spontaneously. We used the template-directed addition of 2-methylimidazole-activated guanylate to an RNA primer as a model system to study the mechanism of polymerization. We found that general acid-base catalysis is not involved and the reaction has a maximum rate at pH 7.5. Inner shell contacts to Mg^{2+} at the reaction center are likely important to orient the 3′-hydroxyl and phosphate for inline attack. Furthermore, we found that there is likely an interaction between activating groups of adjacent monomers, which has important implications for activating group optimization and *in situ* monomer activation.

We then report the synthesis and non-enzymatic polymerization of the non-natural 2′-amino-threose nucleic acids (2′-NH_2-TNA). We found that 2′-NH_2-TNA nucleotides

polymerized more slowly than similar amino-substituted ribonucleic acids. Although single-stranded TNA is less flexible than RNA, we found that polymerization of 2′-NH_2-TNA nucleotides was slower on TNA templates. These results suggest that TNA would have been a poor competitor to RNA in the origin of life.

Finally, we report the development of initiator nucleotides that use 'click' chemistry to site-specifically label RNA transcripts. We propose a scheme for an *in vitro* selection experiment using these nucleotides to isolate a ribozyme that could catalyze RNA replication. Altogether, these results will help guide the development of a self-replicating nucleic acid system that can be integrated with a compartment to create a protocell.

Table of Contents

Acknowledgements

Constructing a complete set of acknowledgements would be a thesis in itself, so many people who helped me get to this point will have to remain unacknowledged. I am indebted to my advisor, Jack Szostak, for assembling a diverse group of talented and creative individuals and giving them the freedom and resources to pursue their ideas. There are surely few more stimulating environments in which to develop as a scientist. Jack is always available to discuss ideas and results, offering incisive critiques and insightful suggestions with disconcerting ease. I am particularly grateful to Alonso Ricardo for taking the time to mentor me as an undergrad and again as a young graduate student. Without his guidance in synthetic chemistry I would likely have fewer fingers. I am also grateful to the undergrads I had the pleasure to mentor, Isabel Vogt and Emily Unger. They brought a lot of energy to the lab and only a small fraction of their work appears in these pages. Ben Heuberger, Matt Powner, Itay Budin, Jesse Chen, Xin Cai, Raphael Bruckner, Na Zhang, Christian Hentrich, Anders Björkbom, Noam Prywes, and Li Li, among many others, have all provided many helpful discussions. My committee members, Gary Ruvkun, Suzanne Walker, and David Liu, provided important guidance and feedback that helped shape the work I will present.

I would not have been able to get through the ups and downs of graduate school without the support of wonderful people outside of lab. My wife Teena has been very supportive and patient with my odd hours, including the occassional stop by lab after an evening out. Finally, I would have never made it here without my parents, Martin and Nancy, my brother Dillon and my sister Lisa. Without such a supportive family I would not have had the confidence to pursue my interests, no matter how foolish.

In memory of my father.

Chapter One: Progress towards synthetic cells

J. Craig Blain and Jack W. Szostak

Note: A version of this chapter will appear in Annual Review of Biochemistry

Abstract

The complexity of even the simplest known life forms makes efforts to synthesize living cells from inanimate components seem like a daunting task. However, recent progress in efforts to create synthetic cells, ranging from simple protocells to artificial cells approaching the complexity of bacteria, suggests that the synthesis of life is now a realistic goal. Protocell research, fueled by advances in the biophysics of primitive membranes and the chemistry of nucleic acid replication, is providing new insights into the origin of cellular life. Parallel efforts to construct more complex artificial cells, incorporating translational machinery and protein enzymes, are providing fresh insights into the requirements for protein-based life. We discuss recent advances and remaining challenges in the synthesis of artificial cells, the possibility of creating new forms of life distinct from existing biology, and the promise of this research for a deeper understanding of the nature of living systems.

1.1 Introduction

Models of the first primitive cells, or protocells, are being created as one way of studying the origin and early evolution of life. Similarly, simplified models of modern cells are being used to test our understanding of the requirements for cellular function, including growth and division. Such divergent goals are inspiring a growing number of laboratories to work towards the synthesis of artificial cells of widely varying composition and complexity. These efforts are united by the goals and challenges of assembling non-living components into living cells.

In this review we will focus on bottom-up approaches to the construction of artificial cells from molecular components or sub-systems (Figure 1.1). Such approaches are well suited to purposes ranging from modeling the earliest life forms, to developing minimal protein-based cells, to exploring the creation of cells based on non-biological components. We begin by considering the genomes of artificial cells, which range from short oligonucleotides for model protocells, to much longer RNAs for artificial cells that incorporate viral polymerases, and on to DNA genomes that are larger still and approach the size of bacterial genomes. We then review recent work on compartments, the supra-molecular structures that define the cellular nature of life. Most compartment boundaries are vesicles composed of bilayer membranes, but these range widely in composition from simple fatty acids to more complex lipids, as well as a variety of synthetic non-biological lipids. Finally, we discuss efforts to integrate replicating genomes with replicating compartments, to generate artificial cells that are capable of sustained reproduction and Darwinian evolution. We also discuss efforts to generate cell-like structures based on non-biological genetic materials and non-biological forms of compartmentalization. We will not

discuss the top-down approach of systematically removing genes from extant organisms to achieve a minimal genome (Hutchison et al. 1999), nor will we consider the creation of new kinds of cells by genome synthesis or extensive genome editing (Gibson et al. 2008; Isaacs et al. 2011).

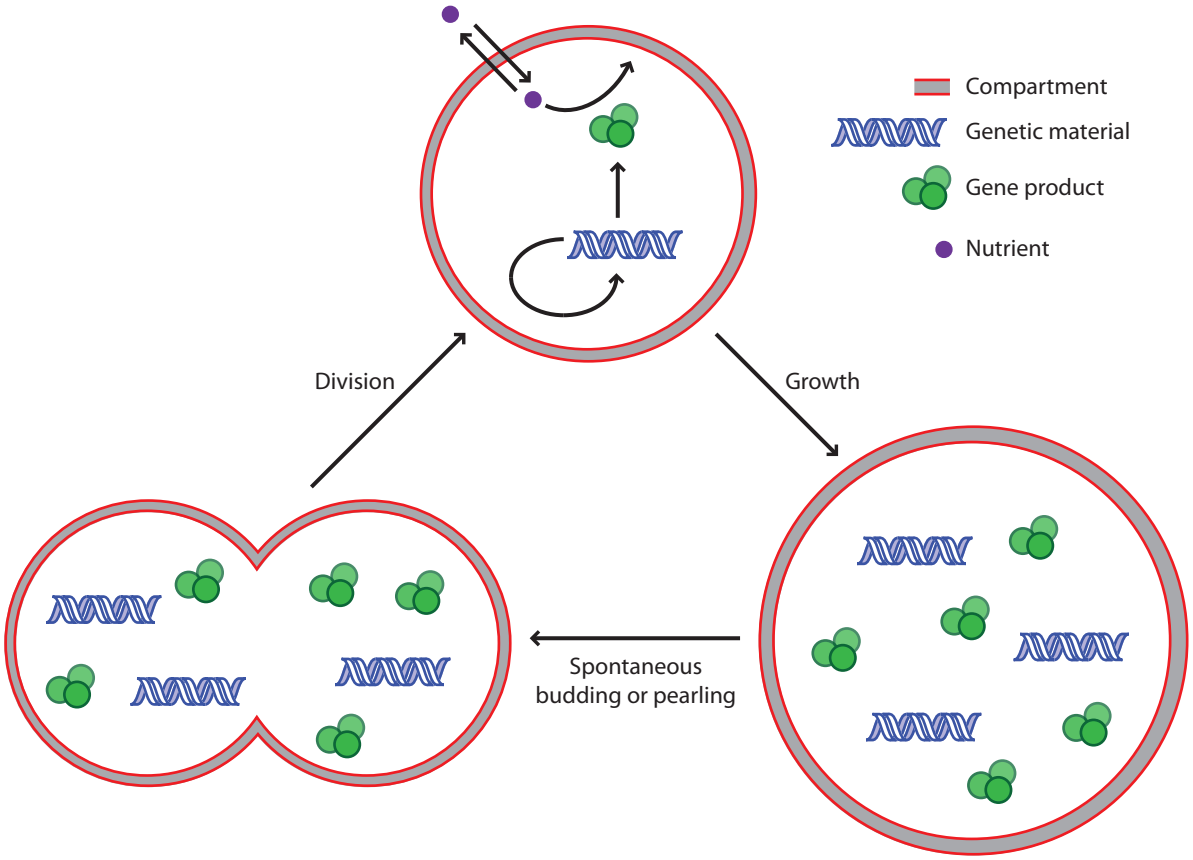


Figure 1.1. A cycle of growth and division for a minimal cell. The genetic material, typically a nucleic acid, acts as a template for the synthesis of gene products, which can be translated proteins or a folded version of the genetic material in the case of ribozyme-based cells. The gene products catalyze replication of the genetic material and carry out basic cellular metabolism, for example, synthesis of compartment components. With highly activated nucleotides, catalysis of genetic replication may not be necessary. Meanwhile, the

Figure 1.1 (continued). compartment, which is typically a phospholipid or fatty acid vesicle, grows through the internal synthesis or external addition of lipids, or by competing with other cells for their lipids. As the cell grows, the system becomes unstable to give the spontaneous budding and division of the cell into small daughter cells that repeat the process.

1.2 Genomes and Genome Replication

To propagate indefinitely a synthetic cell must be able to replicate its own genetic material. In this section we first discuss non-enzymatic and then ribozyme-catalyzed RNA replication, since these are the two proposed processes for genome replication in RNA based protocells. We then discuss the approaches that have been taken for protein-catalyzed genome replication in artificial cells, which typically take advantage of viral RNA or DNA polymerases because of their great simplicity in comparison to cellular DNA replication systems.

1.2.1 Non-enzymatic replication

Nucleic acids are by far the most promising candidates for the genetic material of protocells because of their ability to direct their own replication through template copying (Figure 1.2A). The question of how nucleic acids could replicate before the evolution of genetically encoded polypeptides is central to research on the origin of life (Rich 1962; Woese 1967; Orgel 1968; Crick 1968). After briefly reviewing early work on this topic, we discuss recent advances in non-enzymatic RNA replication, and consider the approaches being taken to overcome the remaining problems with what would appear to be the

simplest form of genetic replication. Finally, we review recent work in which, by dropping the constraints of prebiotic plausibility, chemically modified nucleic acids have been used to achieve rapid copying of mixed sequence templates containing all four bases (Schrum et al. 2009; Zhang et al. 2013; Kaiser and Richert 2013).

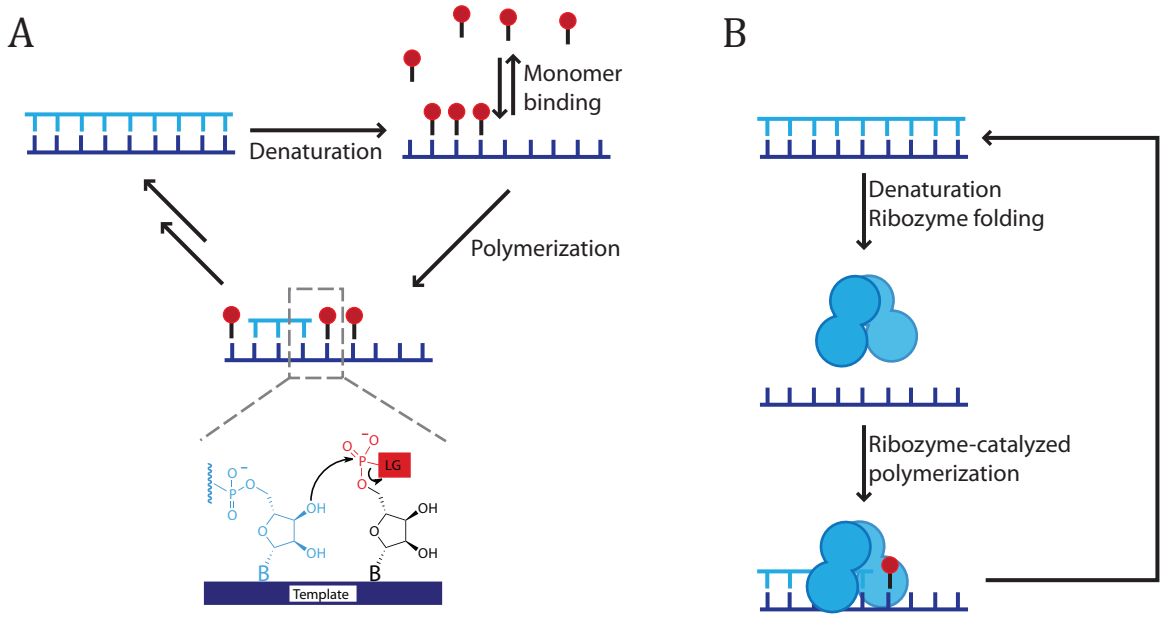


Figure 1.2. Pathways for the self-replication of nucleic acids. (A) In non-enzymatic nucleic acid self-replication activated nucleotides or short oligonucleotides bind to a complementary base in a single-stranded template. The 3′-hydroxyl group of a bound residue will attack the 5′-phosphate of an adjacent residue to displace a leaving group (LG), typically 2-methylimidazole, and form a new phosphodiester bond. This reaction is repeated at each position to copy the entire sequence. Chemical modifications to the natural RNA structure can enhance the rate and fidelity of this process. (B) In ribozyme-catalyzed replication a single-stranded RNA molecule folds into a tertiary structure that can catalyze the template copying reaction. In both models the stability of double-stranded

Figure 1.2 (continued). RNA, hydrolysis RNA and activated nucleotides, and lack of sequence-general copying have limited the potential for self-replication.

The template-enhanced non-enzymatic synthesis of nucleic acids was first demonstrated in the 1960s (Schramm, Grotsch, and Pollman 1962; Naylor and Gilham 1966; Sulston et al. 1968), but the reactions were slow, inefficient and in the case of RNA synthesis led to a predominance of non-natural 2′-5′-linked material. Extensive subsequent work by Leslie Orgel and colleagues showed that ribonucleotides activated with 2-methylimidazole as a leaving group could be used to copy short C-rich RNA templates on a timescale of hours or days without enzymes (Inoue et al. 1984; Haertle and Orgel 1986; Joyce and Orgel 1986; Wu and Orgel 1992a). Although the products were predominantly 3′-5′-linked, the reaction was highly sequence and base-composition dependent, with A and U residues being copied particularly slowly (Wu and Orgel 1992c) and with poor fidelity (Rajamani et al. 2010). Even all GC sequences could not be copied efficiently, for example, the best yield obtained in efforts to copy a 14-mer all GC template was 2%. For these and other reasons, including incomplete regiospecificity, the need for primers, the problem of strand separation, and the lack of mild, specific chemistry for monomer activation (and re-activation following hydrolysis), Orgel and indeed most researchers in the field eventually became convinced that non-enzymatic RNA replication was not chemically realistic (Orgel 2004; Joyce 2002).

Nucleotide activation

The 2′-3′ cyclic phosphate nucleotides, although attractive from the point of view of prebiotic synthesis, are insufficiently activated to lead to efficient template copying (Renz, Lohrmann, and Orgel 1971; Pitsch et al. 1995), while the NTPs used in biology are too unreactive for non-enzymatic polymerization (Rohatgi, Bartel, and Szostak 1996a). Nucleotides activated with 2-methylimidazole (2-MeIm), known as phosphor-2-methylimidazolides, react relatively quickly to preferentially give the natural and more stable (Wasner et al. 1998) 3′-5′ linkage (Inoue and Orgel 1981). Interestingly, other imidazole derivatives do not react as quickly (Inoue and Orgel 1981) and activation with imidazole itself preferentially gives 2′-5′ linkages (Weimann et al. 1968). To date, there is no mechanistic understanding of why 2-methylimidazole activated nucleotides react so quickly and regiospecifically with the 3′-hydroxyl group, although sterics, nucleophile and activating group pK_a's, and interactions between the activating groups of stacked monomers (Wu and Orgel 1992a) may play a role. The rate of addition of guanosine 5′-phosphor-2-methylimidazolide to a primer on a poly(C) template depends on the length of the template, the monomer concentration and the magnesium concentration, but it is typically around one per hour (Prakash, Roberts, and Switzer 1997), whereas the rate of monomer hydrolysis is $\sim 0.02 - 0.05\ h^{-1}$ (Kanavarioti et al. 1989).

Although 2-MeIm is the most widely used activating group, the search for replacements with superior properties continues. Richert and colleagues have made extensive use of 1-hydroxy-7-azabenzotriazole (HOAt), which was originally developed for peptide coupling (Carpino 1993). They have achieved higher rates with this leaving group than 2-MeIm when adding single nucleotides to a 3′-amino modified primer on a DNA

template (Stuetz et al. 2007; Hagenbuch et al. 2005) and in an RNA system (Vogel, Deck, and Richert 2005). It is not yet clear how the two activating groups compare for the addition of multiple nucleotides, or how faster monomer hydrolysis and the requirement for higher pH for HOAt activated nucleotides (Vogel, Deck, and Richert 2005) would affect their utility in a self-replicating system.

Nucleophilic catalysts have been used to generate highly reactive nucleotides *in situ* from 2-MeIm or HOAt activated monomers. In particular, *N*-alkylated imidazole derivatives (Schrum et al. 2009; Zhang et al. 2013) and pyridine (Röthlingshöfer et al. 2008) have been used to accelerate non-enzymatic primer extension reactions with amino-terminated primers, but similar catalysis has not been demonstrated with an RNA system. Adenine derivatives such as 1-methyladenine are also good potential activating groups (Prabahar and Ferris 1997), but their use in template-copying has not been explored. A thorough understanding of the mechanism of non-enzymatic primer extension would allow for the rational design and optimization of improved activating groups, as opposed to the *ad hoc* screening that is currently employed.

Base pairing and fidelity

Polymerases not only accelerate the phosphoryl transfer reaction, they also increase the affinity and specificity of the nucleotide - template interaction (Loeb and Kunkel 1982). Without polymerases, fidelity and sequence context become significant challenges to replication. G and C residues are copied much more quickly than A and U residues due to stronger base-pairing, and the presence of multiple A or U residues in a row in a template drastically reduces copying efficiency (Wu and Orgel 1992c). Furthermore, the comparable

strength of a G-U wobble pair to an A-U base pair (J. L. Chen et al. 2012) leads to significant mis-incorporation of G across from U (Leu et al. 2011).

Two approaches have been taken to overcome these problems. In the first, modified base-pairs are used in place of A-U. For example, the 5-propynyluracil (U^P)-diaminopurine (D) base pair is similar to A-U, but has three hydrogen bonds and improved base stacking. This base-pair has been used to achieve comparable rates of non-enzymatic polymerization to the G-C base pair in multiple systems (J. Chaput, Sinha, and Switzer 2002; Schrum et al. 2009). However, since the propynyl group lowers the pK_a at position N3 of the uracil ring (Budow et al. 2009) it could stabilize the enol tautomer and thus strengthen Watson-Crick-like G:U mis-pairing, similar to the 5-bromo substitution (Katritzky and Waring 1962). Conversely, 2-thio substitution on uracil or thymine strengthens base-pairing to A, but weakens wobble pairing to G due since sulfur is a weak H-bond acceptor (Testa et al. 1999; Sintim and Kool 2006), and may therefore improve both rate and fidelity. Many other alternate base pairs are also possible (Wojciechowski and Leumann 2011). The second approach is to use short oligonucleotides instead of, or in addition to monomers (Szostak 2011; Lohrmann, Bridson, and Orgel 1981) since the cooperativity of base-pairing helps to stabilize individual A-U pairs. High fidelity can be achieved in non-enzymatic oligonucleotide ligation at elevated temperatures (James and Ellington 1997), and oligonucleotide ligation is selective for 3′-5′ linkages (Rohatgi, Bartel, and Szostak 1996b). Activated trimers have been used to demonstrate replication of a hexamer template (Sievers and Kiedrowski 1994), but the poor efficiency of ligation has so far prevented this approach from being extended to longer templates.

After the incorporation of a mismatched nucleotide in a template copying reaction both the rate (Rajamani et al. 2010) and fidelity (Leu et al. 2013) of the next addition decrease. This stalling effect causes accurate template copies to be produced more quickly than mutant sequences, thus increasing the apparent fidelity of the replication process and the length of sequences that can be maintained through replication. Conversely, mutant products tend to incorporate stretches of mismatches and could therefore lead to 'leaps' through sequence space, which could facilitate the evolution of novel structures (Leu et al. 2013).

Backbone modifications

Changing the hydroxyl nucleophile of a ribo- or deoxyribo-nucleotide to a more reactive amine is an effective means of enhancing the rate of non-enzymatic polymerization. Both the 2′-amino and 3′-amino modifications of ribonucleotides were originally explored by Orgel et al. in template copying reactions, and were found to be more reactive than their hydroxyl counterparts (Lohrmann and Orgel 1976; Zielinski and Orgel 1985; Zielinski and Orgel 1987; Tohidi et al. 1987). More recently, our group and that of Clemens Richert have studied the 2′- and 3′-amino-dideoxyribonucleotides in primer extension reactions. Using primers with a 3′-amino terminus and 'helper oligonucleotides' that assist monomer binding, Richert and colleagues have achieved rapid and accurate template-directed single nucleotide primer extension (Kaiser et al. 2012; Kervio et al. 2010). By applying these lessons to an all RNA system, they have been able to efficiently and accurately copy short mixed-sequence RNA templates (Deck, Jauker, and Richert 2011), although this process was quite slow.

We have found that both 2′-amino (Schrum et al. 2009) and 3′-amino (Zhang et al. 2013) modified nucleotides can copy an RNA C_4 template in >80% yield in ~ 10 minutes. The reaction rates are sensitive to the template structure, with A-form helices (RNA and LNA) giving faster rates in both cases than DNA templates. In the 2′-amino system, A-U base pairs were not copied efficiently so the U^P-D pair was used instead. Surprisingly, in the 3′-amino system, both A_4 and T_4 templates were copied efficiently. In a further step towards self-replication, we have used 3′-amino nucleotides to copy mixed sequence templates composed of phosphoramidate DNA (3′-NP-DNA), the product of 3′-amino nucleotide polymerization (Zhang et al.). The G-T wobble base pair gave significant misincorporation in this system, but the use of the 2-thio modification of T, which stabilizes pairing to A and weakens pairing to G (Testa et al. 1999), improved both the fidelity and rate of copying (Zhang et al.). Although 3′-NP-DNA is attractive as a possible genetic polymer for protocells, the copying of longer mixed sequence templates has yet to be demonstrated.

Many modified backbone nucleic acids that are capable of Watson-Crick base-pairing have been synthesized (Hill et al. 2001), but only a few have been examined from the point of view of their ability to support self-replication (Eschenmoser 1999). Threose nucleic acids (TNA) form stable Watson-Crick duplexes with complementary TNA, RNA and DNA strands (Schoning et al. 2000), and have been used as templates in the non-enzymatic polymerization of ribonucleotides (Heuberger and Switzer 2006). Furthermore, an aptamer made of TNA has been isolated by *in vitro* selection (H. Yu, Zhang, and Chaput 2012). Pyranosyl-RNA (p-RNA) has been used as a template for the non-enzymatic ligation of p-RNA oligonucleotides (Pitsch et al. 1995). Interestingly, hexitol nucleic acids (HNA)

(Hendrix et al. 1997) and altritol nucleic acids (ANA) (Allart et al. 1999), which both form single strands that are pre-organized into an A-form geometry, are superior to RNA as templates for the non-enzymatic polymerization of activated ribonucleotides (Kozlov et al. 2000). However, the corresponding activated HNA and ANA monomers do not polymerize efficiently, indicating that subtle differences in monomer structure and reactivity can affect the reaction (Kozlov et al. 2000).

Peptide nucleic acids (PNA) are based on a peptide backbone (Nielsen 1999) but can form Watson-Crick base-pairs with many nucleic acids (Wittung et al. 1994). The peptide backbone means that the whole range of peptide chemistry can be employed in efforts to demonstrate template copying. For example, PNA dimers have been used in DNA template-directed polymerization through EDAC-mediated peptide bond formation (Schmidt et al. 1997), while reductive amination has also proven to be an effective way to polymerize PNA pentamers with aldehyde and amino termini on a DNA template (Kleiner et al. 2008). PNA monomers, however, are prone to cyclization (Schmidt et al. 1997). Conversely, a PNA C_{10} oligomer has been shown to be an effective template for the polymerization of guanosine 5′-phosphorimidazolide (Schmidt, Nielsen, and Orgel 1997). Interestingly, PNA also allows for the template-directed attachment of nucleobase units onto a preformed backbone. This process has been demonstrated using reductive amination to make the process irreversible (Heemstra and Liu 2009), and dynamically using thioester linkages, where the base composition was shown to change in response to changing template sequences (Ura et al. 2009). Since it is uncharged, PNA based on the original aminoethylglycyl backbone is prone to aggregation (Nielsen et al. 1991; Egholm et al. 1992); however, a version of PNA based on a repeating dipeptide motif in which alternating amino acids have charged side chains

(e.g. asp) and nucleobase side chains can form stable, soluble duplexes (Mittapalli et al. 2007).

Other challenges

There are many challenges on the path to non-enzymatic self-replication other than the basic problems of rate and fidelity (Szostak 2012). In particular, strand separation (and thus continuing cycles of replication) can be quite difficult. Even the simplest ribozymes are typically over 30 nucleotides long (McCall et al. 2000); however, an RNA duplex of that length could have a melting temperature over 100°C (Xia et al. 1998), and phosphoramidate DNA is even more stable (Gryaznov et al. 1995). Therefore, in order to replicate functional nucleic acids, thermal denaturation alone will not suffice. Possible solutions to this problem are the use of denaturants such as formamide to destabilize the helix, the replication of short fragments that can assemble into larger functional structures, or the use of heterogeneous backbones. The mixture of 2′-5′ and 3′-5′ linkages that results from non-enzymatic RNA polymerization is usually considered undesirable. However, our laboratory has recently shown that both an aptamer and a ribozyme can maintain function with up to 25% of 2′-5′ linkages randomly interspersed in the sequence (Engelhart, Powner, and Szostak 2013). Since 2′-5′ linkages significantly destabilize a duplex (Wasner et al. 1998) and can be copied non-enzymatically (Prakash, Roberts, and Switzer 1997), they could be beneficial in replicating RNA systems by facilitating strand separation.

Since a continuous supply of activated monomers is required for indefinite self-replication, the tendency of activated monomers to hydrolyze or cyclize is also a serious problem. Furthermore, hydrolysis of activated monomers generates nucleotides that act as

competitive inhibitors of template-directed polymerization (Deck, Jauker, and Richert 2011). The significance of these issues was highlighted by a recent study by Deck *et al.* (Deck, Jauker, and Richert 2011), who were able to efficiently copy an RNA template containing all four natural bases by immobilizing the primer-template complex on magnetic beads and repeatedly exchanging the solution of activated monomers. We have recently obtained a similar improvement in a vesicle-encapsulated RNA copying system, by using dialysis to refresh the external solution with fresh monomers while removing hydrolyzed monomers (Adamala and Szostak). Ultimately a more effective way to solve this problem would be through efficient and selective chemistry for *in situ* re-activation of hydrolyzed monomers. A variety of reagents have been used to activate nucleotides *in situ* including carbodiimides (Sulston et al. 1968), N-cyanoimidazole (Ferris, Huang, and Hagan 1989), and cyanogen bromide (Dolinnaya et al. 1991); however, these reagents also alkylate nucleobases and other nucleophiles (Gilham 1962; Chu, Wahl, and Orgel 1983). A general, mild and efficient means of re-activating monomers in situ would no doubt lead to significant improvements in the efficiency of non-enzymatic RNA replication.

In summary, many of the seemingly insurmountable problems with non-enzymatic nucleic acid replication have been overcome in recent years. While several difficult problems remain, potential solutions to all of these issues have been proposed, such that there is now a sense of optimism that non-enzymatic template-directed replication may be demonstrated in the not to distant future.

1.2.2 Ribozyme-catalyzed replication

Crick, Orgel and Woese first predicted the existence of RNA catalysts, or ribozymes, in the late 1960s, as a solution to the chicken and egg problem of the origin of DNA and protein based life (Woese 1967; Orgel 1968; Crick 1968). The discovery some 15 years later of ribozymes in biology (Kruger et al. 1982; Guerrier-Takada et al. 1983) immediately generated great interest in the possibility of ribozyme-catalyzed RNA replication (Gilbert 1986) (Figure 1.2B). Given substrates with a 5′-terminal guanosine residue, to mimic a splicing intermediate, self-splicing introns were shown to add nucleotides to a single-stranded oligonucleotide (Been and Cech 1988), to ligate multiple oligonucleotides together on a template (Doudna and Szostak 1989), and to extend a primer in a template-dependent manner (Bartel et al. 1991). Indeed, the *sunY* ribozyme can be broken into a multi-subunit complex that can catalyze the template-directed assembly of one of its subunits from a set of oligonucleotides (Doudna, S, and Szostak 1991). Despite the development of a shorter and more active version of the *sunY* ribozyme (Green and Szostak 1992), it was never able to achieve self-replication for at least two reasons: the complementary oligonucleotide substrates bound to and inhibited the ribozyme, and the isoenergetic transesterification reaction used by the ribozyme provided insufficient driving force for the formation of long products.

Due to the inherent limitations of systems based on self-splicing introns, subsequent efforts shifted to the *in vitro* selection of novel ribozymes that could use activated nucleotides to copy templates (Bartel and Szostak 1993). One of the first ribozymes isolated by *in vitro* selection catalyzed the attack of the 3′-hydroxyl of one oligonucleotide on the 5′-triphosphate of a second oligonucleotide, when both were aligned on a template

(Bartel and Szostak 1993). This ribozyme, known as the class I ligase, is able to catalyze the addition of a single nucleotide onto a primer, using an NTP substrate (Ekland and Bartel 1996). McGinness and Joyce (K. McGinness, Wright, and Joyce 2002) used continuous *in vitro* evolution to select for a variant of this ribozyme that can catalyze three successive nucleotidyl transfer reactions in both 5′→3′ and 3′→5′ directions. However, the polymerase activity conferred by the evolved point substitutions was modest, and it was the evolution of a novel accessory domain that first transformed the class I ligase into a true polymerase. The resulting R18 ribozyme can catalyze the template-directed elongation of a primer by up to 14 nucleotides, again using NTPs as substrates (Johnston et al. 2001). The R18 ribozyme is a proof-of-principle that ribozymes can catalyze RNA polymerization, an assumption at the heart of the RNA world hypothesis. It has also been a useful starting point for further *in vitro* selection experiments with the ultimate goal of achieving RNA self-replication (Zaher and Unrau 2007).

To date, the best RNA polymerase ribozyme is the tC19Z variant of the R18 ribozyme (Wochner et al. 2011), developed through a combination of *in vitro* selection and engineering. It contains several base substitutions that improve activity and, critically, was engineered to bind the 5′-end of its substrate through sequence complementarity to a region of the ribozyme. The R18 ribozyme has very low affinity for its primer-template substrate, but catalyzes the reaction rapidly and with some processivity once bound (Lawrence and Bartel 2003). By building in a sequence-specific interaction with the template, Wochner *et al.* observed a dramatic increase in ribozyme activity. On a template composed of repeats of an optimized 11 nt sequence, the tC19Z ribozyme is able to extend a primer by up to 91 nt (albeit with a yield of only 0.035%). The mutation rate, as assessed

by sequencing full-length products, was 8.8×10^{-3}, although this value is likely an underestimate since it does not include products that may have stalled after a mis-incorporation event (Rajamani et al. 2010). This ribozyme polymerase was used to synthesize a catalytically active 24 nt portion of the mini-Hammerhead ribozyme, the first example of a ribozyme being synthesized by another ribozyme. Although this was a major advance, R18 is close to 200 nt in length, so a significant increase in activity, processivity and sequence generality would be required to even approach the copying of its own sequence. Furthermore, R18 is highly structured, with many self-complementary regions that are strong blocks to continued template copying. Finally, R18 is only active in the presence of high concentrations of Mg^{2+}, which catalyzes RNA hydrolysis, including degradation of the ribozyme itself.

Further efforts to optimize R18-derived ribozymes may be facilitated by the recently determined structure of the class I ligase core (Shechner et al. 2009; Shechner and Bartel 2011). Globally, the class I ligase resembles a tripod with three helical legs converging on the active site (Shechner et al. 2009). Similarly to most proteinaceous polymerases, the class I ligase uses an Mg^{2+} ion to coordinate the 3′-hydroxyl group of the primer and α-phosphate; however, it does not use the second ion of the canonical two metal-ion mechanism. Instead, the pyrophosphate leaving group is oriented back into the major groove and nucleobase and hydroxyl groups are thought to stabilize the transition state electrostatically through hydrogen bonding. The accessory domain is less well characterized, but Wang *et al.* (Wang, Cheng, and Unrau 2011) used mutational analysis to identify critical secondary structures and tertiary interactions. They suggested that the domain is likely draped over the vertex of the class I ligase tripod structure.

Since the point mutations identified by selection experiments on the R18 ribozyme have resulted in only modest improvements, it is possible that the ribozyme represents a local maximum of activity in sequence space. If that is the case, the significant advance required to achieve self-replication may have to come either from the evolution of one or more additional accessory domains, or from a different polymerase core entirely. After developing the R18 polymerase, Lawrence and Bartel (Lawrence and Bartel 2005) selected for eight other accessory domains for the class I ligase. Although none of the new ribozymes were better than the R18 polymerase, they offer different starting points for further selection. Accessory domains conferring enhanced processivity would seem to offer the greatest potential for improvement. Similarly, a number of independent core catalytic domains are now available, and some have been transformed into nascent polymerases. McGinness and Joyce (K. E. McGinness and Joyce 2002) evolved the hc ligase into the 18-2 ribozyme which can catalyze the addition of single nucleotides to a primer using NTPs as substrates. Other natural (Vicens and Cech 2009) and selected (Rogers and Joyce 2001) ligase ribozymes have the potential to be evolved into polymerases; however, the direct selection for polymerase ribozymes with high affinity for the primer-template substrate would avoid the difficulty of evolving an RNA that was initially selected to perform one function into a variant with a different function.

Polymerase activity has the advantage of allowing for an open-ended exploration of sequence space; however, systems with more limited evolutionary potential have been used to explore aspects of self-replication in an RNA system. Joyce and colleagues (Paul and Joyce 2002; Kim and Joyce 2004; Lincoln and Joyce 2009) developed a system based on the R3C ligase ribozyme that is able to ligate together two halves of another R3C ligase (Rogers

and Joyce 2001), which can then ligate together two halves of the original ligase ribozyme. This system shows the exponential growth characteristic of self-replication (Ferretti and Joyce 2013). Limited evolutionary dynamics were incorporated into this system by coupling the particular sequence of a variable four nucleotide 'genotype' in the substrate recognition arm to a particular sequence of a variable region in the ligase active site (Sczepanski and Joyce 2012). After continuous *in vitro* evolution, the population was dominated by sequences with a more active ligase 'phenotype' that preferentially replicated ligases with the same genotype, as well as inactive parasitic sequences that were recognized by the active ligases through stable mispairing at the genotype site. However, the diversity of this system is inherently limited by the length of the genotype and phenotype sequences, which must be provided externally.

Another approach to ribozyme self-replication that is based on fragment ligation, but in this case using splicing chemistry, has been developed by Lehman and colleagues (Riley and Lehman 2003; Hayden and Lehman 2006; Hayden, Kiedrowski, and Lehman 2008; Vaidya et al. 2012). This three-ribozyme system was able to amplify itself through a cooperative network of cross-catalysis. Although the compartmentalization of such self-replicating systems could in principle be coupled to cell level evolutionary selection (I. Chen, Roberts, and Szostak 2004), achieving continued replication inside protocells would be difficult because of the need to feed the system with large oligonucleotide substrates, while retaining ligated products in the cell. Furthermore, the evolutionary capabilities of such systems are inherently limited by the large, defined RNA substrates that must be provided, whereas polymerase based replication systems can more readily evolve novel functions through open-ended exploration of sequence space.

1.2.3 Protein-catalyzed replication

When proteins are introduced to the system, the issue of nucleic acid genome replication becomes relatively simple. Many enzyme-based methods have been used to replicate nucleic acid sequences, including the polymerase chain reaction (PCR) (Oberholzer, Albrizio, and Luisi 1995), the Qβ replicase (Oberholzer et al. 1995), and reverse transcriptase/RNA polymerase systems (Guatelli et al. 1990; Wright and Joyce 1997).

PCR has been used to model genome replication in several artificial cell systems. Although most polymerases are strongly inhibited by fatty acids, PCR proceeds well in phospholipid vesicles (Oberholzer, Albrizio, and Luisi 1995; Kurihara et al. 2011). Deamer *et al*. (Chakrabarti et al. 1994) found that although phospholipid membranes are normally impermeable to NTP substrates, temperature cycling across the lipid melting transition creates defects which allow NTPs to enter vesicles, thus allowing for continued RNA replication. Phospholipid vesicles are also quite robust to the temperature fluctuations required for PCR. Sugawara *et al*. have used PCR to replicate DNA within more complex replicating vesicles composed of phospholipids and cationic lipids (Kurihara et al. 2011).

The Qβ replicase system is particularly interesting since it uses the RNA-dependent RNA-polymerase from the Qβ bacteriophage, which can replicate its genomic RNA isothermally without the use of primers (Haruna and Spiegelman 1965). This system was the first used to demonstrate the evolution of a nucleic acid sequence *in vitro* (Mills, Peterson, and Spiegelman 1967) and has been successfully incorporated into vesicles (Oberholzer et al. 1995). Yomo *et al*. have extensively studied the activity of Qβ replicase in

model protocells (Kita et al. 2008; Ichihashi et al. 2008; Urabe et al. 2010; Bansho et al. 2012).

With protein-based replication, the problem of genome replication is deferred to the problem of replicating the protein polymerase, which necessitates incorporation of the complex translational machinery. *In vitro* translation systems have successfully been incorporated into vesicles, resulting in the efficient synthesis of proteins (Oberholzer, Nierhaus, and Luisi 1999; Noireaux and Libchaber 2004), including the Qβ replicase itself (Kita et al. 2008). However, replicating the entire system including DNA and RNA polymerases, transcription and translation factors, tRNAs, the ribosome, and other components is still a distant goal and proposals to do so involve over 100 genes (Forster and Church 2006; Luisi, Ferri, and Stano 2005). Properly coordinating the replication of such a large synthetic system may require additional regulatory components.

1.3 Compartments and Compartment Replication

Work on developing self-replicating compartments for model protocells is relatively advanced, in that complete replication cycles of growth and division have been demonstrated. Fatty acid based vesicles are prebiotically plausible, and have many physical properties that are appropriate for a protocell with minimal (or no) evolved biochemical machinery. For more advanced artificial cell models, more complex membranes are required. Phospholipid based membranes seem, so far at least, to be essential for avoiding polymerase inhibition by fatty acids; however, the use of phospholipid membranes raises a host of problems, from the need to import NTPs and other small molecule substrates through these more cohesive membranes, to the need for mechanisms to drive membrane

growth and division. Generally these problems are solved through the use of additional proteins, further increasing the complexity of the system, although in some cases novel physical processes can provide alternative solutions. Alternative forms of encapsulation or segregation, such as emulsion (Figure 1.3A), coacervate and ATPS systems are also being explored.

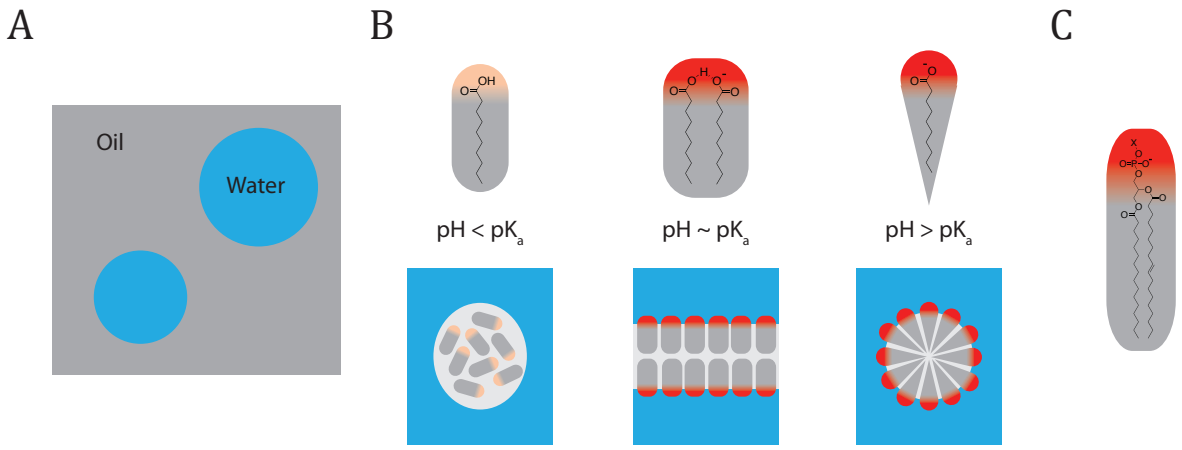


Figure 1.3. Common components of protocell compartments. (A) Nucleic acids, proteins and other hydrophilic species can be compartmentalized in water-in-oil emulsions by stabilizing the water droplets with surfactants. Droplets can fuse and divide by vigorous agitation. (B) Fatty acids spontaneously aggregate into larger structures above a certain concentration. At a pH below their pK_a of ~7-9 these aggregates are amorphous precipitates, and above their pK_a repulsion of the charged carboxylate groups leads to the formation of small micelles. However, near their pK_a the fatty acids are partially ionized and form a hydrogen-bonding network that stabilizes the formation of extended bilayers. These bilayers can curve onto themselves to form enclosed vesicles. (C) Phospholipids have

Figure 1.3 (continued). two fatty acid chains per charged head group and therefore form stable bilayers under a wider range of conditions than fatty acids.

1.3.1 Fatty acid basic vesicles

The protocell membrane is a compartment boundary that serves two critical functions. First, it provides a selective barrier between the cell and its environment, allowing the influx of nutrients and the efflux of waste while stably encapsulating the macromolecular contents of the protocell. Second, it provides the physical link between functional gene products and the genome from which they were expressed - this physical link is necessary for Darwinian evolutionary selection. Compartmentalization also allows for the selection of cooperative gene networks. Without compartmentalization, self-replicating systems are prone to parasitism from species that are replicated efficiently (Szathmary and Smith 1995). For a protocell membrane to be viable as part of a model for the origin of cellular life, it should be able to assemble, grow and divide spontaneously without the assistance of any gene products.

Fatty acids form bilayer membranes when the solution pH is approximately the pKa of the carboxylate head groups in the membrane, allowing them to form transient hydrogen-bonded dimers (Haines 1983; Cistola et al. 1988) (Figure 1.3B). The formation of vesicles at this pH is highly cooperative and is characterized by a critical aggregation concentration (cac) of fatty acid monomers above which vesicles spontaneously self-assemble (Tanford 1980). Unlike phospholipids, fatty acids rapidly exchange between vesicle membranes, micelles and dissolved monomers. It is this dynamic exchange that makes fatty acids particularly well suited as components of protocell membranes, because

such vesicles can grow either through the external addition of fatty acids (I. A. Chen and Szostak 2004), or by absorbing material from other vesicles (I. Chen, Roberts, and Szostak 2004; Budin and Szostak 2011), as well as through the generation of fatty acids *in situ* from precursors (Walde et al. 1994). Furthermore, multilamellar vesicles (*i.e.*, vesicles with multiple membranes) form thread-like shapes as they grow that readily divide into spherical daughter vesicles with the application of shear forces (Zhu and Szostak 2009) or following photochemically induced pearling (Zhu et al. 2012). These cycles of growth and division can be repeated without the loss of vesicle contents (Zhu and Szostak 2009). We have recently shown that this process can be driven simply by the concentration of a lipid solution by solvent evaporation (Budin, Debnath, and Szostak 2012) (Figure 1.4). Division of spherical vesicles can be induced by extrusion through narrow pores, although this process results in partial loss of contents (Hanczyc, Fujikawa, and Szostak 2003).

Fatty acid membranes are much more permeable than phospholipid membranes to small charged molecules. In particular, activated nucleotides can pass through the membrane and take part in template-directed polymerization inside the protocell to produce oligonucleotide products that remain trapped inside (Mansy et al. 2008). To a certain extent, the permeability (Sacerdote and Szostak 2005; Mansy et al. 2008), stability (Mansy and Szostak 2008; Monnard et al. 2002), and dynamics (Budin and Szostak 2011; Bruckner et al. 2009; Apel, Deamer, and Mautner 2002) of fatty acid vesicles can be tuned by changing the acyl-chain length and saturation state of the fatty acid components, or through the addition of additives such as fatty alcohols, fatty esters and polycyclic aromatic hydrocarbons.

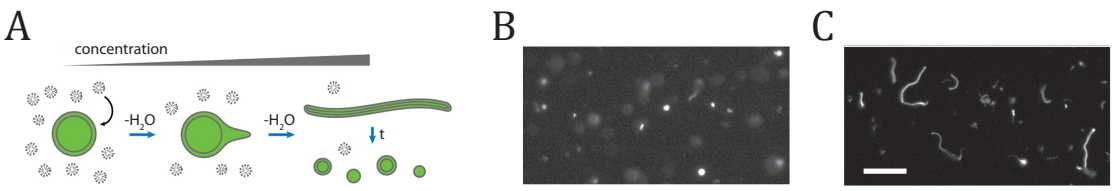


Figure 1.4. Demonstration of fatty acid vesicle growth and division. Work from our group has shown that concentrating multilamellar fatty acid vesicles by evaporating the water solvent causes the membrane to grow into long tubules, which can divide into small vesicles through simple agitation. (A) Schematic diagram of the growth and division process. (B) Fluorescence microscropy image of vesicles before concentration. (C) Fluorescence microscopy image of vesicles after concentrations. The contents of the vesicle are maintained during this growth and division process. Similar processes may have been an important mechanism of replication for protocells on the early earth. Adapted with permission from Budin I, Debnath A, and Szostak JW. 2012. Concentration-Drive Growth of Model Protocell Membranes. *J. Am. Chem. Soc.* 134(51):20812-20819. Copyright 2012 American Chemical Society.

Fatty acids are particularly interesting due to their likely involvement in the origin of life. They were almost certainly available on the early earth since they have been found in meteorites, albeit in low levels (Yuen and Kvenvolden 1973; Naraoka, Shimoyama, and Harada 1999), they can be synthesized abiotically (Rushdi and Simoneit 2001) and they readily self-assemble into membranes under potentially prebiotic conditions (Deamer and Barchfield 1982; Hanczyc, Fujikawa, and Szostak 2003). Vesicle dynamics may have provided an early opportunity for selection based on competition for fatty acids (I. Chen, Roberts, and Szostak 2004; Budin and Szostak 2011). However, vesicles composed of fatty

acids are not as robust as those made of phospholipids. In particular, they precipitate in the presence of millimolar concentrations of divalent cations, which are required for non-enzymatic RNA polymerization (Monnard et al. 2002) and many enzymes.

1.3.2 Phospholipid and Synthetic Lipid Vesicles

Phospholipid vesicles are the compartment of choice for more complex artificial cells; their biophysical properties have been studied extensively (Tanford 1980) (Figure 1.3C). They are stable to a wide range of conditions and they can be formed by a variety of different techniques (Jesorka and Orwar 2008). However, because phospholipid molecules are essentially permanently anchored within membranes, and therefore do not exchange between membranes, the growth of phospholipid vesicles must occur through different processes than those that lead to the growth of fatty acid vesicles. Under certain conditions phospholipid vesicles can grow through fusion with other vesicles (Müller et al. 2003; Terasawa et al. 2012; Johnson et al. 2002; Sunami, Caschera, et al. 2010), but vesicle-vesicle fusion mediated by fusogenic lipids or peptides is often associated with significant contents leakage. In principle, *in situ* enzymatic lipid synthesis (Kuruma et al. 2009; Murtas 2010) should allow unlimited growth as long as appropriate substrates are supplied, and the necessary enzymes are generated internally by translation. Division by budding can be induced by the depletion volume effect (Terasawa et al. 2012) or phase separation (Andes-Koback and Keating 2011) of encapsulated polymers, or through enzymatic activity (Staneva, Angelova, and Koumanov 2004; Takakura and Sugawara 2004).

Membrane vesicle growth and division using non-biological lipids and small molecule catalysts has been explored by Sugawara *et al*. (Takakura, Toyota, and Sugawara

2003; Takahashi et al. 2009; Kurihara et al. 2011). In one system, an amphipilic aldehyde is generated *in situ* through the action of a membrane-localized catalyst on a protected precursor; the aldehyde then condenses with an amphilic amine to form a cationic bola-amphiphile (Kurihara et al. 2011). In related experiments, the *in situ* synthesis of membrane forming amphiphiles was found to perturb the structure of multilamellar vesicles so as to cause either vesicle division, or the assembly and release of new vesicles from the interior of a pre-formed vesicle. The combination of catalyzed lipid synthesis and subsequent vesicle division is an exciting advance towards self-replicating systems of compartments that are chemically distinct from anything seen in biology.

1.3.3 Alternative approaches to compartmentalization

Although phospholipid and fatty acid vesicles are the most studied models for the membranes of artificial cells, other systems have been developed. As reviewed by Kamat and Hammer (Kamat, Katz, and Hammer 2011), a variety of polymers have been used to create vesicles known as polymersomes, including polyethylene derivatives (Discher et al. 1999), polypeptides (Holowka, Pochan, and Deming 2005), and dendrimers (Percec et al. 2010). Polymersomes can be readily engineered to have particular desirable properties and functionality. Because their stability makes growth and division difficult (Robbins et al. 2009; Mabrouk et al. 2009; Kamat et al. 2010) artificial cells based on polymersomes are typically used to model aspects of non-growing cells. The design of subunits that would generate polymersomes with the dynamic properties required for growth (either continuous or via fusion events) and division is an interesting challenge for future research. Moving even further away from standard biology, artificial cells based on non-membrane

types of compartmentalization have recently been studied. Water-in-oil emulsions have been used to encapsulate the Qβ replicase system (Urabe et al. 2010) and complete transcription-translation systems (Tawfik and Griffiths 1998; Pietrini and Luisi 2004). Remarkably, the aqueous droplets of emulsions can be 'fed' by fusion with other droplets, and can be forced to divide by shear forces or extrusion, allowing for a cycle of growth and division (Ichihashi et al.). The eutectic phase channels formed in frozen water have been used to spatially partition a ribozyme polymerase (Attwater et al. 2010) and non-enzymatic RNA polymerization reactions (Monnard, Kanavarioti, and Deamer 2003), although such systems are not amenable to cycles of replication.

Interestingly, the components of a protocell can also be localized through selective partitioning in multi-phase systems. Functional ribozymes can be enriched in the dextran-rich droplets that form spontaneously in aqueous mixtures of PEG and dextran (Strulson et al. 2012). Nucleotides and cationic peptides can assemble into coacervate microdroplets that can sequester enzymes and other components (Koga et al. 2011). An important question is whether the lack of a low permeability membrane barrier in these systems results in rapid exchange of RNAs (or other genetic materials) between droplets, which could diminish or abolish the spatial partitioning of genetic materials that is required for Darwinian evolution. The combination of ATPS or coacervate systems with lipid membranes provides the opportunity for compartmentalization within a protocell, analogous to that obtained in biological cells by organelles; furthermore, such combined systems provide a novel pathway for protocell division in response to osmotic and surface tension effects (Andes-Koback and Keating 2011).

1.4 Integrated Artificial Cellular Systems

Interesting new challenges and opportunities arise when self-replicating genetic systems and compartments are brought together. Here we discuss the distinct issues arise in the design of integrated protocell models, vs. more complex protein-based artificial cells.

1.4.1 Simple protocells

In an attempt to resolve the 'genetics first' vs. 'compartments first' controversy in the origin of life field, one of us (JWS) together with Pier Luigi Luisi and David Bartel (Szostak, Bartel, and Luisi 2001) proposed an integrated protocell model based on a self-replicating nucleic acid genome within a self-replicating membrane boundary. That model was based on ribozyme-catalyzed RNA replication, since at the time there was no plausible path towards non-enzymatic RNA replication. The model also proposed that the emergence of a second ribozyme that carried out some cellular function, e.g. synthesis of a lipid component of a vesicle compartment, would result in an integrated cell capable of Darwinian evolution. Since the proposal of that early model, advances in both nucleic acid and compartment replication have brought the field to the point where a variety of protocell models are being constructed and evaluated.

A fundamental aspect of any viable protocell model is that the physico-chemical conditions required for stability and replication of the nucleic acid and membrane components must be compatible. Fatty acid vesicles are stable at a pH from about 7 to 9 depending on the particular fatty acid used (Cistola et al. 1988; Haines 1983; Fukuda et al. 2001), and this can be extended to a wider range by the incorporation of fatty alcohols and glycerol monoesters (Maurer et al. 2009). Fortunately, this pH range is favorable for both

non-enzymatic RNA polymerization (Wu and Orgel 1992b; Deck, Jauker, and Richert 2011) and RNA polymerase ribozyme activity (Wochner et al. 2011; Johnston et al. 2001). Fatty acid based vesicles can also be stable up to 100 ^{o}C, thus allowing for thermal denaturation of encapsulated nucleic acids, as well as enhanced permeation of nucleotide substrates (Mansy and Szostak 2008). Unfortunately, fatty acid membranes are unstable above low millimolar concentrations of divalent cations (Monnard et al. 2002), whereas non-enzymatic RNA polymerization (Wu and Orgel 1992b) and ribozyme polymerase activity (Johnston et al. 2001) typically require over 50 mM Mg^{2+}.

There are several potential solutions to this apparent incompatibility. Non-enzymatic template-directed copying with amino-sugar nucleotides does not require Mg^{2+} and has successfully been carried out in fatty acid vesicles (Mansy et al. 2008), suggesting that artificial cells based on phosphoramidate nucleic acids may be possible. However, a protocell with an RNA genome and ribozyme catalysts would require alternative solutions - either a replacement for divalent cations that does not destroy fatty acid membranes, or a means of complexing the essential Mg^{2+} ion so as to protect the membrane while allowing RNA replication to proceed. We have recently found that the tricarboxylic acid citrate chelates Mg^{2+} such that fatty acid membranes are preserved, while non-enzymatic template copying is only minimally affected (Adamala and Szostak). This has allowed us to encapsulate an RNA primer-template complex inside fatty acid vesicles, add activated ribonucleotides to the vesicles, and observe primer-extension as the nucleotides spontaneously diffuse across the membrane and then copy the template strands inside the vesicle. This advance is sufficient to allow the further development of model protocells based on RNA, but it also raises the exciting prospect that chelators with catalytic activity

could facilitate the copying of long mixed-sequence templates. An alternative approach to the compatibility problem would be the isolation of a ribozyme replicase that functions at very low levels of Mg^{2+} ions, which may be feasible since biological ribozymes function at low intracellular Mg^{2+} levels. Finally, it may be possible to replace the fatty acids of the protocell membranes with alternative lipids such as nonionic amphiphiles that do not interact strongly with Mg^{2+} ions.

Another requirement for a fully integrated protocell is that nutrients, in particular activated nucleotides, must be able to enter the cell, and waste must be able to leave, but the self-replicating genetic system must be permanently trapped inside. Fatty acid vesicles have an advantage in this case since they are permeable to small charged molecules, including nucleotides, but not oligonucleotides (Mansy and Szostak 2008). Raising the temperature or adding low millimolar concentrations of Mg^{2+} further increases their permeability (Mansy and Szostak 2008). Phospholipid vesicles are impermeable to even monovalent cations; however, small charged molecules are able to travel through transient defects in the bilayer membrane (Paula et al. 1996) that form at the gel to liquid phase transition temperature (Kanehisa and Tsong 1978). This property has been exploited to allow for feeding of external NTPs to encapsulated enzymatic reactions (Chakrabarti et al. 1994), although this method may require thermocycling between the membrane phase transition temperature and the temperature optimum of the enzyme (Monnard, Luptak, and Deamer 2007). Alternatively, protein pores (Noireaux and Libchaber 2004; Yoshimoto et al. 2005), shorter-chain phospholipids (Nourian, Roelofsen, and Danelon 2012), or detergents (Treyer, Walde, and Oberholzer 2002) can be used to permeabilize the membrane. Another attractive approach is to use vesicle fusion to deliver nutrients

encapsulated in feeder vesicles (Sunami, Caschera, et al. 2010); unfortunately, most simple vesicle fusion methods are inefficient and also induce considerable contents leakage (Müller et al. 2003; Terasawa et al. 2012; Johnson et al. 2002; Sunami, Hosoda, et al. 2010).

By directly coupling the protocell genome to cell-level phenotypes, compartmentalization could drive the evolution of more complex functions in a way that could emulate the early evolution of life. Perhaps the simplest such coupling results from the osmotic pressure due to encapsulated RNA (I. Chen, Roberts, and Szostak 2004). The membranes of osmotically swollen vesicles are under tension, but this high energy state can relax through the absorption of fatty acids from surrounding vesicles that are less swollen because they contain less RNA. Thus, any mutations that enhance RNA replication, leading to the faster accumulation of internal RNA, will lead to faster membrane growth. While osmotically driven competitive growth is attractive because of the simplicity of the physically mediated linkage between genome replication and cell growth, osmotically swollen vesicles are very difficult to divide.

An alternative, but still quite simple, means of linking genomic function to cell growth would involve a genomically encoded ribozyme with phospholipid synthase activity. We have recently shown that fatty acid vesicles that contain even small amounts of phospholipids (Budin and Szostak 2011) will grow at the expense of surrounding vesicles with less phospholipid content. This effect is independent of osmotic pressure, resulting instead from the ordering effect of phospholipids on the membrane, which causes a decreased rate of dissociation of fatty acid molecules from the membrane. The net effect is that genomic sequences leading to a low level of phospholipid synthesis (e.g. by condensation of a lysophospholipid with a fatty acyl thioester) would confer a strong

growth advantage on the protocell; moreover, the resulting growth follows the filamentous growth pathway, so that subsequent division is facile. A very similar growth advantage results from the internal synthesis of hydrophobic peptides (Adamala and Szostak 2013), which could also be catalyzed either directly by a ribozyme, or indirectly by a catalytic peptide that is in turn synthesized by a ribozyme. These processes would lead to a strong selective advantage for genomically encoded and thus heritable phospholipid or peptide synthesis activity, respectively, given the availability of appropriate substrates.

For any artificial cellular system to be capable of indefinite propagation, the replication of the genome, biochemical contents and compartment must be coordinated. Interestingly, this may be less of a problem in very simple protocells, than in more complex artificial cells that contain many components. In a simple protocell that uses either non-enzymatic or ribozyme catalyzed genome replication, with post-synthesis strand separation driven by high temperature excursions, genome replication would be self-limiting since strand reannealing is a second order process. Thus, increasing concentrations of genomic RNA would lead to faster reannealing, until reannealing was faster than strand copying - as a result, a steady state level of genomic RNA would be attained. On the other hand, if compartment replication is much faster than genome replication the population could be overrun by empty protocells. While this could be avoided by the controlled addition of limiting quantities of micelles to fatty acid vesicles (Zhu and Szostak 2009), if growth is driven by an internal process such as ribozyme-promoted phospholipid synthesis (Budin and Szostak 2011), then excess cell membrane growth would lead to ribozyme dilution, which would decrease the rate of ribozyme-promoted growth. Thus the combination of ribozyme-promoted growth and concentration-

dependent strand reannealing could, in principle, lead to a steady state level of genomic RNA in a population of replicating protocells, without additional regulatory signals.

An interesting version of a similar process at the DNA level was recently demonstrated by Sugawara *et al*. (Kurihara et al. 2011) who showed that PCR mediated DNA amplification could drive the growth and division of cationic vesicles. In this system, the vesicles replicate through the adsorption of a cationic lipid precursor, which is hydrolyzed by a catalyst to generate the final cationic membrane component. Although the mechanism of the DNA enhanced growth and division is not entirely clear, it is likely a consequence of the interaction of the anionic DNA with the cationic membrane, leading to enhanced adsorption of the cationic precursor, and possibly DNA induced curvature changes that lead to enhanced vesicle division.

1.4.2 Complex artificial cells

Since many of the applications proposed for artificial cells (Forster and Church 2006; Pohorille and Deamer 2002) require proteins, much work has been done to develop systems for efficient compartmentalized transcription and translation. Early experiments by Oberholzer *et al.* (Oberholzer, Nierhaus, and Luisi 1999) showed that polyuridylic acid could be translated into polyphenylalanine in phospholipid vesicles. Yu *et al.* (W. Yu et al. 2001) and Nomura *et al.* (Nomura et al. 2003) were able to synthesize green fluorescent protein (GFP) in greater yield using a cell-free *E. coli* extract encapsulated in giant phospholipid vesicles (>1 μm in diameter), but yields were still limited by the impermeability of the phospholipid membrane, which made it impossible to replenish nucleotides and amino acids. Noireaux *et al.* (Noireaux and Libchaber 2004) solved the

problem of vesicle impermeability by expressing the α-hemolysin pore protein in their vesicles, which allowed them to maintain translation for up to four days by adding nutrients externally. In a further step towards a totally synthetic cell, Murtas *et al.* (Murtas et al. 2007) incorporated the PURE translation system (Shimizu et al. 2001), which is reconstituted from purified components, into phospholipid vesicles. Nevertheless, even this system has a limited life time, as critical components degrade or become inactivated. One of the ultimate goals of this line of research is to have the encapsulated transcription-translation machinery regenerate all of the cell components, thus prolonging synthetic activity and ultimately leading to a complete self-replicating system - in effect the reconstitution or semi-synthesis of a simplified bacterial cell. However, there are many challenges that must be addressed before this goal can be approached.

For complex artificial cells containing elaborate biochemical networks with larger numbers of components, the problem of maintaining the proper balance between replication of the genome, the biochemical machinery, and the compartment itself, becomes more complex. Even in relatively simple systems, regulation can quickly become challenging. For example, in a recent study the Qβ replicase was translated from Qβ genomic RNA, which was concurrently replicated by the Qβ replicase (Kita et al. 2008). In this system, the ribosome and the Qβ replicase inhibited each other's activity by competing for binding to the Qβ RNA. As a result, efficient genome amplification could only be obtained at an optimized ratio of ribosomes to replicase (Ichihashi et al. 2008).

When an artificial cell divides its contents will randomly partition to the two daughter cells; the more components the system has the more copies will be required to ensure that both daughter cells acquire at least one copy of each essential component. Thus

it may become important to design in simple regulatory feedback mechanisms to keep multi-component systems such as the translation apparatus internally balanced during growth. Reliance on statistical segregation of many components may impose a minimum cell size, such that all components can be synthesized at a high copy number prior to cell division. Alternatively, designing in mechanisms for the non-random segregation of key components could increase the efficiency of cell division.

A very interesting way to avoid this problem has recently been demonstrated by Yomo *et al.*, who developed an emulsion droplet model of a replicating artificial cell (Ichihashi et al.). In this system, the droplets contain a transcription-translation system. Droplet growth is controlled by fusion with smaller droplets containing fresh transcription-translation mixture, while division is controlled by periodic passage of the emulsion through a filter with pores of a defined size. The cycle of growth and division is thus manually controlled, but can be continued indefinitely, and indeed the system was taken through almost 200 generations of growth and division. In a remarkable example of the spontaneous emergence of Darwinian behavior, a genomic RNA coding for the Qß replicase was encapsulated, with the RNA being replicated by its encoded replicase. Over many generations, the RNA and the protein replicase co-evolved enhanced binding affinity and specificity, leading to more efficient synthesis of the replicase and thus more efficient replication of the RNA. This system is in principle open-ended, and could be used to evolve additional RNA-encoded proteins, as long as a means of imposing an appropriate selective pressure can be found. The demonstration of Darwinian evolution in this droplet replication system is quite impressive, and makes this clearly the most advanced artificial cell model yet.

Although gene expression in phospholipid vesicles is now routine, efforts to translate proteins that function in the growth or division of the artificial cell are still in the early stages. A step in this direction is the expression in vesicles of two enzymes from a pathway that produces phospholipids from glycerol phosphate and fatty acyl-CoA (Kuruma et al. 2009). Although the expected lipid products were detectable, the efficiency of the synthesis was far too low to result in measurable membrane growth. In a step towards internal energy metabolism, a single subunit of the F_1F_0-ATPase has been expressed in vesicles containing the remaining subunits to complete the formation of a functional ATPase complex (Kuruma et al. 2012). Although very crude cell division may not require any cytoskeletal machinery, as evidenced by the growth of L-form mutant bacteria (Leaver et al. 2009; Allan, Hoischen, and Gumpert 2009), more controlled cell division requires the actin and tubulin homologs mreB and ftsZ. The expression of the actin homolog mreB in liposomes results in the assembly of fibril bundles on the membrane surface, but only if the membrane protein mreC is co-expressed. Interestingly, the fibril bundles can perturb the shape of the liposome, suggesting that effects on division may not be too difficult to reconstitute. Additional pathways that must be fully reconstituted to enable an artificial cell to replicate include the replication of a moderately large DNA genome of roughly 100 kB, and the assembly of functional ribosomes.

1.5 Future of the Field

For nucleic acid-based protocells, the development of a self-replicating genome remains the most significant challenge. Enhancing non-enzymatic template copying through chemical modifications (Schrum et al. 2009; Zhang et al. 2013; Kaiser and Richert

2013) and/or the use of a mixture of activated mononucleotides and short oligonucleotides (Szostak 2011), and developing novel RNA-dependent RNA polymerase ribozymes are promising avenues of research. Once such a replication system is developed, it should become possible to construct integrated replicating protocells, and then to follow the emergence of Darwinian evolution in a protocell population seeded with random-sequence templates. The selection of metabolic ribozymes with useful activities such as phospholipid synthesis (Budin and Szostak 2011), or the evolution of structural RNAs that modulate cell division are both interesting possibilities, but perhaps the most interesting possibility would be the spontaneous emergence of completely unexpected function. Protocells should be especially good at selecting for complexes or networks of ribozymes that function together since their relationship will be maintained as they replicate - such experiments are essentially impossible in the absence of compartmentalization. Once any RNA that confers an advantage on its host cell emerges, there should be strong selection for the emergence of ribozymes that enhance replication efficiency and accuracy. Such experiments may eventually shed light on the emergence of metabolism and translation, and may address the question of the cellular complexity attainable in an RNA based cell. Protocell research has already contributed to the development of plausible pathways and geochemical scenarios for the origin of life (Ricardo and Szostak 2009; Szostak 2012), and an improved understanding of the requirements for protocell reproduction should further constrain such models. Nucleic acid-based protocells are generally considered to have fewer practical applications than those capable of expressing proteins. However, given the wide variety of aptamers and ribozymes that have been isolated (Lee et al. 2004) and the rapid developments in the fields of biosensing (J. Liu, Cao, and Lu 2009), DNA

nanotechnology (Ke et al. 2012) and computation (Qian and Winfree 2011), the functional potential of nucleic acids should not be underestimated.

For artificial cellular systems based on internal protein expression an entirely different set of challenges lies ahead. In essence, all pathways required for the growth and division of a minimal bacterial cell must be fully reconstituted so as to function with an efficiency approaching that seen in extant biological cells. Thus DNA replication, transcription, translation, ribosome assembly and membrane synthesis must all be reconstituted and built into the design of the artificial cell. In the process of building and combining these substrates, many interesting questions can be addressed. For example, can all of metabolism be avoided simply by incorporating membrane channels and supplying nutrients externally? With respect to self-assembly, are there components other than the genome and the membrane that can only be made in an auto-catalytic or self-templated manner? Candidates for such materials include the peptidoglycan cell wall which grows by addition of new material to old, or membrane protein chaperones that are required for their own folding (C. L. Hagan, Westwood, and Kahne 2013). Only the complete reconstitution of a living cell from purified components will prove that all such 'self-templating' materials and processes have been identified. The synthesis of a living artificial cell from components will open the door to many even more adventurous lines of research, e.g. the synthesis of 'enantio-life' in which all components have reversed chirality (Bohannon 2010), and the synthesis of cells with further reduced complexity, approximating the nature of intermediates in the evolutionary path from protocells to modern life. It may even become possible to assemble cells in which the standard biopolymers are replaced with altered versions, as a means of exploring the range of

molecular diversity that is compatible with cellular life. Finally, there is no shortage of proposed applications for artificial cells (Leduc et al. 2007; Forster and Church 2006; Pohorille and Deamer 2002), ranging from the biosynthesis of pharmaceuticals (Ro et al. 2006) or biofuels (Rabinovitch-Deere et al. 2013) to advanced drug delivery technology in which artificial cells would be engineered to control the timing and site of drug release. Artificial cells might be designed to continuously synthesize and deliver a drug within a human host (Leduc et al. 2007), signal to and modify the behavior of natural cells, or to carry out functions usually performed by healthy human cells, for example the transport of oxygen through the blood (Chang 1999). The exciting challenge of the synthesis of life from non-living components, combined with the potential for transformative applications, will no doubt drive progress in this field for many years to come.

1.6 References

Adamala, K, and J W Szostak. "Non-Enzymatic Template-Directed RNA Synthesis Inside Model Protocells." *Science* in press.

Adamala, Katarzyna, and J W Szostak. 2013. "Competition Between Model Protocells Driven by an Encapsulated Catalyst." *Nature Chemistry* 5 (6) (June): 495–501. doi:10.1038/nchem.1650.

Allan, E J, C Hoischen, and J Gumpert. 2009. "Bacterial L-Forms." *Advances in Applied Microbiology* 68: 1–39. doi:10.1016/S0065-2164(09)01201-5.

Allart, B, K Khan, H Rosemeyer, G Schepers, C Hendrix, K Rothenbacher, F Seela, A Van Aerschot, and P Herdewijn. 1999. "D-Altritol Nucleic Acids (ANA): Hybridisation Properties, Stability, and Initial Structural Analysis." *Chemistry-a European Journal* 5 (July 20): 2424–2431.

Andes-Koback, Meghan, and Christine D Keating. 2011. "Complete Budding and Asymmetric Division of Primitive Model Cells to Produce Daughter Vesicles with Different Interior and Membrane Compositions." *Journal of the American Chemical Society* 133 (24) (June 22): 9545–9555. doi:10.1021/ja202406v.

Apel, Charles L, D W Deamer, and Michael N Mautner. 2002. “Self-Assembled Vesicles of Monocarboxylic Acids and Alcohols: Conditions for Stability and for the Encapsulation of Biopolymers.” *Biochimica Et Biophysica Acta* 1559 (1) (February 10): 1–9.

Attwater, James, Aniela Wochner, Vitor B Pinheiro, Alan Coulson, and Philipp Holliger. 2010. “Ice as a Protocellular Medium for RNA Replication.” *Nature Communications* 1: 76. doi:10.1038/ncomms1076.

Bansho, Yohsuke, Norikazu Ichihashi, Yasuaki Kazuta, Tomoaki Matsuura, Hiroaki Suzuki, and Tetsuya Yomo. 2012. “Importance of Parasite RNA Species Repression for Prolonged Translation-Coupled RNA Self-Replication.” *Chemistry & Biology* 19 (4) (April 20): 478–487. doi:10.1016/j.chembiol.2012.01.019.

Bartel, D P, and J W Szostak. 1993. “Isolation of New Ribozymes From a Large Pool of Random Sequences.” *Science* 261 (5127) (September 10): 1411–1418.

Bartel, D P, J A Doudna, N Usman, and J W Szostak. 1991. “Template-Directed Primer Extension Catalyzed by the Tetrahymena Ribozyme.” *Molecular and Cellular Biology* 11 (6): 3390–3394.

Been, M D, and T R Cech. 1988. “RNA as an RNA Polymerase: Net Elongation of an RNA Primer Catalyzed by the Tetrahymena Ribozyme.” *Science* 239 (4846): 1412–1416.

Bohannon, J. 2010. “MIrror-Image Cells Could Transform Science -- or Kill Us All.” *Wired*, November 29.

Bruckner, R J, S S Mansy, A Ricardo, L Mahadevan, and J W Szostak. 2009. “Flip-Flop-Induced Relaxation of Bending Energy: Implications for Membrane Remodeling.” *Biophysical Journal* 97 (12) (December 16): 3113–3122. doi:10.1016/j.bpj.2009.09.025.

Budin, Itay, and J W Szostak. 2011. “Physical Effects Underlying the Transition From Primitive to Modern Cell Membranes.” *Proceedings of the National Academy of Sciences USA* 108 (13) (March 29): 5249–5254. doi:10.1073/pnas.1100498108.

Budin, Itay, Anik Debnath, and J W Szostak. 2012. “Concentration-Driven Growth of Model Protocell Membranes.” *Journal of the American Chemical Society* 134 (51) (December 26): 20812–20819. doi:10.1021/ja310382d.

Budow, Simone, Henning Eickmeier, Hans Reuter, and Frank Seela. 2009. “2'-Deoxy-5-Propynyluridine: a Nucleoside with Two Conformations in the Asymmetric Unit.” *Acta Crystallographica. Section C, Crystal Structure Communications* 65 (Pt 12) (December): o645–8. doi:10.1107/S0108270109044850.

Carpino, L A. 1993. “1-Hydroxy-7-Azabenzotriazole - an Efficient Peptide Coupling Additive.” *Journal of the American Chemical Society* 115 (10): 4397–4398.

Chakrabarti, A C, R R Breaker, G F Joyce, and D W Deamer. 1994. “Production of RNA by a

Polymerase Protein Encapsulated Within Phospholipid Vesicles." *Journal of Molecular Evolution* 39 (6) (December): 555–559.

Chang, T M. 1999. "Future Prospects for Artificial Blood." *Trends in Biotechnology* 17 (2) (February): 61–67.

Chaput, JC, S Sinha, and C Switzer. 2002. "5-Propynyluracil-Diaminopurine: an Efficient Base-Pair for Non-Enzymatic Transcription of DNA." *Chemical Communications* (15): 1568–1569. doi:10.1039/b204535d.

Chen, IA, RW Roberts, and J W Szostak. 2004. "The Emergence of Competition Between Model Protocells." *Science* 305 (5689): 1474–1476.

Chen, Irene A, and J W Szostak. 2004. "A Kinetic Study of the Growth of Fatty Acid Vesicles." *Biophysical Journal* 87 (2) (August): 988–998. doi:10.1529/biophysj.104.039875.

Chen, Jonathan L, Abigael L Dishler, Scott D Kennedy, Ilyas Yildirim, Biao Liu, Douglas H Turner, and Martin J Serra. 2012. "Testing the Nearest Neighbor Model for Canonical RNA Base Pairs: Revision of GU Parameters." *Biochemistry* 51 (16) (April 24): 3508–3522. doi:10.1021/bi3002709.

Chu, B C, G M Wahl, and L E Orgel. 1983. "Derivatization of Unprotected Polynucleotides." *Nucleic Acids Research* 11 (18) (September 24): 6513–6529.

Cistola, D P, J A Hamilton, D Jackson, and D M Small. 1988. "Ionization and Phase Behavior of Fatty Acids in Water: Application of the Gibbs Phase Rule." *Biochemistry* 27 (6) (March 22): 1881–1888.

Crick, F H C. 1968. "The Origin of the Genetic Code." *Journal of Molecular Biology* 38 (3): 367–379.

Deamer, D W, and G L Barchfield. 1982. "Encapsulation of Macromolecules by Lipid Vesicles Under Simulated Prebiotic Conditions." *Journal of Molecular Evolution* 18 (June 4): 203–206.

Deck, Christopher, Mario Jauker, and Clemens Richert. 2011. "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA." *Nature Chemistry* 3 (8) (August): 603–608. doi:10.1038/nchem.1086.

Discher, B M, Y Y Won, D S Ege, J C Lee, F S Bates, D E Discher, and D A Hammer. 1999. "Polymersomes: Tough Vesicles Made From Diblock Copolymers." *Science* 284 (5417) (May 14): 1143–1146.

Dolinnaya, N G, N I Sokolova, D T Ashirbekova, and Z A Shabarova. 1991. "The Use of BrCN for Assembling Modified DNA Duplexes and DNA-RNA Hybrids; Comparison with Water-Soluble Carbodiimide." *Nucleic Acids Research* 19 (11) (June 11): 3067–3072.

Doudna, JA, and J W Szostak. 1989. “RNA-Catalyzed Synthesis of Complementary-Strand RNA.” *Nature* 339 (6225): 519–522.

Doudna, JA, Couture S, and J W Szostak. 1991. “A Multisubunit Ribozyme That Is a Catalyst of and Template for Complementary Strand RNA Synthesis.” *Science* 251 (5001): 1605–1608.

Egholm, M, O Buchardt, P E Nielsen, and R H Berg. 1992. “Peptide Nucleic Acids (PNA). Oligonucleotide Analogues with an Achiral Peptide Backbone.” *Journal of the American Chemical Society* 114 (May 19): 1895–1897.

Ekland, EH, and D P Bartel. 1996. “RNA-Catalysed RNA Polymerization Using Nucleoside Triphosphates.” *Nature* 382 (6589): 373–376.

Engelhart, Aaron E, Matthew W Powner, and J W Szostak. 2013. “Functional RNAs Exhibit Tolerance for Non-Heritable 2′-5′ Versus 3′-5′ Backbone Heterogeneity.” *Nature Chemistry* 5 (5) (May): 390–394. doi:10.1038/nchem.1623.

Eschenmoser, A. 1999. “Chemical Etiology of Nucleic Acid Structure.” *Science* 284 (5423) (June 25): 2118–2124.

Ferretti, Antonio C, and Gerald F Joyce. 2013. “Kinetic Properties of an RNA Enzyme That Undergoes Self-Sustained Exponential Amplification.” *Biochemistry* 52 (7) (February 19): 1227–1235. doi:10.1021/bi301646n.

Ferris, J P, C H Huang, and W J Hagan. 1989. “N-Cyanoimidazole and Diimidazole Imine: Water-Soluble Condensing Agents for the Formation of the Phosphodiester Bond.” *Nucleosides & Nucleotides* 8 (3): 407–414. doi:10.1080/07328318908054184.

Forster, Anthony C, and George M Church. 2006. “Towards Synthesis of a Minimal Cell.” *Molecular Systems Biology* 2: 45. doi:10.1038/msb4100090.

Fukuda, Hiroshi, Ayako Goto, Hisashi Yoshioka, Rensuke Goto, Kenichi Morigaki, and Peter Walde. 2001. “Electron Spin Resonance Study of the pH-Induced Transformation of Micelles to Vesicles in an Aqueous Oleic Acid/Oleate System.” *Langmuir* 17 (14) (July): 4223–4231. doi:10.1021/la0100338.

Gibson, Daniel G, Gwynedd A Benders, Cynthia Andrews-Pfannkoch, Evgeniya A Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B Stockwell, et al. 2008. “Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma Genitalium Genome.” *Science* 319 (5867): 1215–1220. doi:10.1126/science.1151721.

Gilbert, W. 1986. “The RNA World.” *Nature* 319 (6055): 618–618.

Gilham, P T. 1962. “An Addition Reaction Specific for Uridine and Guanosine Nucleotides and Its Application to the Modification of Ribonuclease Action.” *Journal of the American Chemical Society* 84 (October 17): 687–688.

Green, R, and J W Szostak. 1992. “Selection of a Ribozyme That Functions as a Superior Template in a Self-Copying Reaction.” *Science* 258 (5090) (December 18): 1910–1915.

Gryaznov, S, D H Lloyd, J K Chen, R G Schultz, L A DeDionisio, L Ratmeyer, and W D Wilson. 1995. “Oligonucleotide N3′-P5′ Phosphoramidates.” *Proceedings of the National Academy of Sciences USA* 92 (13): 5798–5802.

Guatelli, J C, K M Whitfield, D Y Kwoh, K J Barringer, D D Richman, and T R Gingeras. 1990. “Isothermal, in Vitro Amplification of Nucleic Acids by a Multienzyme Reaction Modeled After Retroviral Replication.” *Proceedings of the National Academy of Sciences USA* 87 (5) (March): 1874–1878.

Guerrier-Takada, C, K Gardiner, T Marsh, N Pace, and S Altman. 1983. “The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme.” *Cell* 35 (3): 849–857.

Haertle, T, and L E Orgel. 1986. “Template-Directed Synthesis on the Oligonucleotide D(C7-G-C7).” *Journal of Molecular Biology* 188 (1) (March 5): 77–80.

Hagan, Christine L, David B Westwood, and Daniel Kahne. 2013. “Bam Lipoproteins Assemble BamA in Vitro.” *Biochemistry* 52 (August 21): 6108–6113. doi:10.1021/bi400865z.

Hagenbuch, Patrizia, Eric Kervio, Annette Hochgesand, Ulrich Plutowski, and Clemens Richert. 2005. “Chemical Primer Extension: Efficiently Determining Single Nucleotides in DNA.” *Angewandte Chemie-International Edition in English* 44 (40) (October 14): 6588–6592. doi:10.1002/anie.200501794.

Haines, T H. 1983. “Anionic Lipid Headgroups as a Proton-Conducting Pathway Along the Surface of Membranes: a Hypothesis.” *Proceedings of the National Academy of Sciences USA* 80 (1) (January): 160–164.

Hanczyc, Martin M, Shelly M Fujikawa, and J W Szostak. 2003. “Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division.” *Science* 302 (5645) (October 24): 618–622. doi:10.1126/science.1089904.

Haruna, I, and S Spiegelman. 1965. “Autocatalytic Synthesis of a Viral RNA in Vitro.” *Science* 150 (3698) (November 12): 884–886.

Hayden, Eric J, and Niles Lehman. 2006. “Self-Assembly of a Group I Intron From Inactive Oligonucleotide Fragments.” *Chemistry & Biology* 13 (8) (August): 909–918. doi:10.1016/j.chembiol.2006.06.014.

Hayden, Eric J, Günter Von Kiedrowski, and Niles Lehman. 2008. “Systems Chemistry on Ribozyme Self-Construction: Evidence for Anabolic Autocatalysis in a Recombination Network.” *Angewandte Chemie-International Edition in English* 47 (44) (October 20): 8424–8428. doi:10.1002/anie.v47:44.

Heemstra, Jennifer M, and David R Liu. 2009. “Templated Synthesis of Peptide Nucleic Acids via Sequence-Selective Base-Filling Reactions.” *Journal of the American Chemical Society* 131 (32): 11347–11349. doi:10.1021/ja904712t.

Hendrix, C, H Rosemeyer, B De Bouvere, A Van Aerschot, F Seela, and P Herdewijn. 1997. “1′,5′-Anhydrohexitol Oligonucleotides: Hybridisation and Strand Displacement with Oligoribonucleotides, Interaction with RNase H and HIV Reverse Transcriptase.” *Chemistry-a European Journal* 3 (July 28): 1513–1520.

Heuberger, Benjamin D, and Christopher Switzer. 2006. “Nonenzymatic Oligomerization of RNA by TNA Templates.” *Organic Letters* 8 (25) (December): 5809–5811. doi:10.1021/ol062368s.

Hill, DJ, MJ Mio, RB Prince, TS Hughes, and JS Moore. 2001. “A Field Guide to Foldamers.” *Chemical Reviews* 101 (12): 3893–4011. doi:10.1021/cr990120t.

Holowka, Eric P, Darrin J Pochan, and Timothy J Deming. 2005. “Charged Polypeptide Vesicles with Controllable Diameter.” *Journal of the American Chemical Society* 127 (35) (September 7): 12423–12428. doi:10.1021/ja053557t.

Hutchison, CA, SN Peterson, SR Gill, RT Cline, O White, CM Fraser, HO Smith, and JC Venter. 1999. “Global Transposon Mutagenesis and a Minimal Mycoplasma Genome.” *Science* 286 (5447): 2165–2169.

Ichihashi, N, K Usui, Y Kazuta, T Sunami, T Matsuura, and T Yomo. “Darwinian Evolution of a Translation-Coupled RNA Self-Replication System in a Cell-Like Compartment.” *Nature Communications* in press.

Ichihashi, Norikazu, Tomoaki Matsuura, Hiroshi Kita, Kazufumi Hosoda, Takeshi Sunami, Koji Tsukada, and Tetsuya Yomo. 2008. “Importance of Translation-Replication Balance for Efficient Replication by the Self-Encoded Replicase.” *Chembiochem* 9 (18) (December 15): 3023–3028. doi:10.1002/cbic.200800518.

Inoue, T, and L E Orgel. 1981. “Substituent Control of the PolyC-Directed Oligomerization of Guanosine 5'-Phosphorimidazolide.” *Journal of the American Chemical Society* 103 (25): 7666–7667.

Inoue, T, G F Joyce, K Grzeskowiak, L E Orgel, J M Brown, and C B Reese. 1984. “Template-Directed Synthesis on the Pentanucleotide CpCpGpCpC.” *Journal of Molecular Biology* 178 (3) (September 25): 669–676.

Isaacs, Farren J, Peter A Carr, Harris H Wang, Marc J Lajoie, Bram Sterling, Laurens Kraal, Andrew C Tolonen, et al. 2011. “Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement.” *Science* 333 (6040) (July 15): 348–353. doi:10.1126/science.1205822.

James, K D, and A D Ellington. 1997. “Surprising Fidelity of Template-Directed Chemical

Ligation of Oligonucleotides." *Chemistry & Biology* 4 (8) (August): 595–605.

Jesorka, Aldo, and Owe Orwar. 2008. "Liposomes: Technologies and Analytical Applications." *Annual Review of Analytical Chemistry* 1: 801–832. doi:10.1146/annurev.anchem.1.031207.112747.

Johnson, Joseph M, Taekjip Ha, Steve Chu, and Steven G Boxer. 2002. "Early Steps of Supported Bilayer Formation Probed by Single Vesicle Fluorescence Assays." *Biophysical Journal* 83 (6) (December): 3371–3379. doi:10.1016/S0006-3495(02)75337-X.

Johnston, WK, PJ Unrau, MS Lawrence, ME Glasner, and D P Bartel. 2001. "RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension." *Science* 292 (5520): 1319–1325.

Joyce, G F, and L E Orgel. 1986. "Non-Enzymic Template-Directed Synthesis on RNA Random Copolymers. Poly(C, G) Templates." *Journal of Molecular Biology* 188 (3) (April 5): 433–441.

Joyce, GF. 2002. "The Antiquity of RNA-Based Evolution." *Nature* 418 (6894): 214–221. doi:10.1038/418214a.

Kaiser, Andreas, and Clemens Richert. 2013. "Nucleotide-Based Copying of Nucleic Acid Sequences Without Enzymes." *The Journal of Organic Chemistry* 78 (3) (February 1): 793–799. doi:10.1021/jo3025779.

Kaiser, Andreas, Sebastian Spies, Tanja Lommel, and Clemens Richert. 2012. "Template-Directed Synthesis in 3′- and 5′-Direction with Reversible Termination." *Angewandte Chemie-International Edition in English* 51 (33) (August 13): 8299–8303. doi:10.1002/anie.201203859.

Kamat, Neha P, Gregory P Robbins, Jeffrey S Rawson, Michael J Therien, Ivan J Dmochowski, and Daniel A Hammer. 2010. "A Generalized System for Photo-Responsive Membrane Rupture in Polymersomes." *Advanced Functional Materials* 20 (16) (August 23): 2588–2596. doi:10.1002/adfm.201000659.

Kamat, Neha P, Joshua S Katz, and Daniel A Hammer. 2011. "Engineering Polymersome Protocells." *The Journal of Physical Chemistry Letters* 2 (13) (July 7): 1612–1623. doi:10.1021/jz200640x.

Kanavarioti, A, CF Bernasconi, D L Doodokyan, and D J Alberas. 1989. "Magnesium-Ion Catalyzed P-N Bond Hydrolysis in Imidazolide-Activated Nucleotides - Relevance to Template-Directed Synthesis of Polynucleotides." *Journal of the American Chemical Society* 111 (18): 7247–7257.

Kanehisa, M I, and T Y Tsong. 1978. "Cluster Model of Lipid Phase Transitions with Application to Passive Permeation of Molecules and Structure Relaxations in Lipid

Bilayers." *Journal of the American Chemical Society* 100 (2) (September 5): 424–432.

Katritzky, A R, and A J Waring. 1962. "Tautomeric Azines. Part I. the Tautomerism of 1-Methyluracil and 5-Bromo-1-Methyluracil." *Journal of the Chemical Society*: 1540–1544. doi:10.1039/jr9620001540.

Ke, Yonggang, Luvena L Ong, William M Shih, and Peng Yin. 2012. "Three-Dimensional Structures Self-Assembled From DNA Bricks." *Science* 338 (6111) (November 30): 1177–1183. doi:10.1126/science.1227268.

Kervio, Eric, Annette Hochgesand, Ulrich E Steiner, and Clemens Richert. 2010. "Templating Efficiency of Naked DNA." *Proceedings of the National Academy of Sciences USA* 107 (27) (July 6): 12074–12079. doi:10.1073/pnas.0914872107.

Kim, Dong-Eun, and Gerald F Joyce. 2004. "Cross-Catalytic Replication of an RNA Ligase Ribozyme." *Chemistry & Biology* 11 (11) (November): 1505–1512. doi:10.1016/j.chembiol.2004.08.021.

Kita, Hiroshi, Tomoaki Matsuura, Takeshi Sunami, Kazufumi Hosoda, Norikazu Ichihashi, Koji Tsukada, Itaru Urabe, and Tetsuya Yomo. 2008. "Replication of Genetic Information with Self-Encoded Replicase in Liposomes." *Chembiochem* 9 (15) (October 13): 2403–2410. doi:10.1002/cbic.200800360.

Kleiner, Ralph E, Yevgeny Brudno, Michael E Birnbaum, and David R Liu. 2008. "DNA-Templated Polymerization of Side-Chain-Functionalized Peptide Nucleic Acid Aldehydes." *Journal of the American Chemical Society* 130 (14): 4646–4659. doi:10.1021/ja0753997.

Koga, Shogo, David S Williams, Adam W Perriman, and Stephen Mann. 2011. "Peptide-Nucleotide Microdroplets as a Step Towards a Membrane-Free Protocell Model." *Nature Chemistry* 3 (9) (August 7): 720–724. doi:10.1038/nchem.1110.

Kozlov, IA, M Zielinski, B Allart, L Kerremans, A Van Aerschot, R Busson, P Herdewijn, and L E Orgel. 2000. "Nonenzymatic Template-Directed Reactions on Altritol Oligomers, Preorganized Analogues of Oligonucleotides." *Chemistry-a European Journal* 6 (1): 151–155.

Kruger, K, P J Grabowski, A J Zaug, J Sands, D E Gottschling, and Cech T R. 1982. "Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena." *Cell* 31 (1): 147–157.

Kurihara, Kensuke, Mieko Tamura, Koh-ichiroh Shohda, Taro Toyota, Kentaro Suzuki, and Tadashi Sugawara. 2011. "Self-Reproduction of Supramolecular Giant Vesicles Combined with the Amplification of Encapsulated DNA." *Nature Chemistry* 3 (10) (October): 775–781. doi:10.1038/nchem.1127.

Kuruma, Yutetsu, Pasquale Stano, Takuya Ueda, and Pier Luigi Luisi. 2009. "A Synthetic

Biology Approach to the Construction of Membrane Proteins in Semi-Synthetic Minimal Cells." *Biochimica Et Biophysica Acta* 1788 (2) (February): 567–574. doi:10.1016/j.bbamem.2008.10.017.

Kuruma, Yutetsu, Toshiharu Suzuki, Sakurako Ono, Masasuke Yoshida, and Takuya Ueda. 2012. "Functional Analysis of Membranous F O- Asubunit of F 1F O-ATP Synthase by in Vitroprotein Synthesis." *Biochemical Journal* 442 (3) (February 24): 631–638. doi:10.1093/bioinformatics/btm404.

Lawrence, MS, and D P Bartel. 2003. "Processivity of Ribozyme-Catalyzed RNA Polymerization." *Biochemistry* 42 (29): 8748–8755. doi:10.1021/bi0342281.

Lawrence, MS, and D P Bartel. 2005. "New Ligase-Derived RNA Polymerase Ribozymes." *Rna* 11 (8): 1173–1180. doi:10.1261/rna.2110905.

Leaver, M, P Domínguez-Cuevas, J M Coxhead, R A Daniel, and J Errington. 2009. "Life Without a Wall or Division Machine inBacillus Subtilis." *Nature* 457 (7231) (February 12): 849–853. doi:10.1038/nature07742.

Leduc, Philip R, Michael S Wong, Placid M Ferreira, Richard E Groff, Kiryn Haslinger, Michael P Koonce, Woo Y Lee, et al. 2007. "Towards an in Vivo Biologically Inspired Nanofactory." *Nature Nanotechnology* 2 (1) (January): 3–7. doi:10.1038/nnano.2006.180.

Lee, Jennifer F, Jay R Hesselberth, Lauren Ancel Meyers, and Andrew D Ellington. 2004. "Aptamer Database." *Nucleic Acids Research* 32 (Database issue) (January 1): D95–100. doi:10.1093/nar/gkh094.

Leu, Kevin, Benedikt Obermayer, Sudha Rajamani, Ulrich Gerland, and Irene A Chen. 2011. "The Prebiotic Evolutionary Advantage of Transferring Genetic Information From RNA to DNA." *Nucleic Acids Research* 39 (18) (October): 8135–8147. doi:10.1093/nar/gkr525.

Leu, Kevin, Eric Kervio, Benedikt Obermayer, Rebecca M Turk-MacLeod, Caterina Yuan, Jesus-Mario Luevano, Eric Chen, Ulrich Gerland, Clemens Richert, and Irene A Chen. 2013. "Cascade of Reduced Speed and Accuracy After Errors in Enzyme-Free Copying of Nucleic Acid Sequences." *Journal of the American Chemical Society* 135 (1) (January 9): 354–366. doi:10.1021/ja3095558.

Lincoln, Tracey A, and Gerald F Joyce. 2009. "Self-Sustained Replication of an RNA Enzyme." *Science* 323 (5918) (February 27): 1229–1232. doi:10.1126/science.1167856.

Liu, Juewen, Zehui Cao, and Yi Lu. 2009. "Functional Nucleic Acid Sensors." *Chemical Reviews* 109 (5) (May 13): 1948–1998. doi:10.1021/cr030183i.

Loeb, L A, and T A Kunkel. 1982. "Fidelity of DNA Synthesis." *Annual Review of Biochemistry* 51: 429–457. doi:10.1146/annurev.bi.51.070182.002241.

Lohrmann, R, and L E Orgel. 1976. “Template-Directed Synthesis of High Molecular Weight Polynucleotide Analogues.” *Nature* 261 (5558) (May 27): 342–344.

Lohrmann, R, P K Bridson, and L E Orgel. 1981. “Condensation of Activated Diguanylates on a Poly(C) Template.” *Journal of Molecular Evolution* 17 (5): 303–306.

Luisi, Pier Luigi, Francesca Ferri, and Pasquale Stano. 2005. “Approaches to Semi-Synthetic Minimal Cells: a Review.” *Naturwissenschaften* 93 (1) (November 15): 1–13. doi:10.1007/s00114-005-0056-z.

Mabrouk, Elyes, Damien Cuvelier, Françoise Brochard-Wyart, Pierre Nassoy, and Min-Hui Li. 2009. “Bursting of Sensitive Polymersomes Induced by Curling.” *Proceedings of the National Academy of Sciences USA* 106 (18) (April 28): 7294–7298. doi:10.1073/pnas.0813157106.

Mansy, Sheref S, and J W Szostak. 2008. “Thermostability of Model Protocell Membranes.” *Proceedings of the National Academy of Sciences USA* 105 (36) (September 9): 13351–13355. doi:10.1073/pnas.0805086105.

Mansy, Sheref S, Jason P Schrum, Mathangi Krishnamurthy, Sylvia Tobé, Douglas A Treco, and J W Szostak. 2008. “Template-Directed Synthesis of a Genetic Polymer in a Model Protocell.” *Nature* 454 (7200) (July 3): 122–125. doi:10.1038/nature07018.

Maurer, S E, D W Deamer, J M Boncella, and P A Monnard. 2009. “Chemical Evolution of Amphiphiles: Glycerol Monoacyl Derivatives Stabilize Plausible Prebiotic Membranes.” *Astrobiology* 9 (10) (December): 979–987. doi:10.1089/ast.2009.0384.

McCall, M J, P Hendry, A A Mir, J Conaty, G Brown, and T J Lockett. 2000. “Small, Efficient Hammerhead Ribozymes.” *Molecular Biotechnology* 14 (1) (January): 5–17. doi:10.1385/MB:14:1:5.

McGinness, Kathleen E, and Gerald F Joyce. 2002. “RNA-Catalyzed RNA Ligation on an External RNA Template.” *Chemistry & Biology* 9 (3) (March): 297–307.

McGinness, KE, MC Wright, and GF Joyce. 2002. “Continuous in Vitro Evolution of a Ribozyme That Catalyzes Three Successive Nucleotidyl Addition Reactions.” *Chemistry & Biology* 9 (5): 585–596.

Mills, DR, RL Peterson, and S Spiegelman. 1967. “An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule.” *Proceedings of the National Academy of Sciences USA* 58 (1): 217–224.

Mittapalli, Gopi Kumar, Kondreddi Ravinder Reddy, Hui Xiong, Omar Munoz, Bo Han, Francesco De Riccardis, Ramanarayanan Krishnamurthy, and Albert Eschenmoser. 2007. “Mapping the Landscape of Potentially Primordial Informational Oligomers: Oligodipeptides and Oligodipeptoids Tagged with Triazines as Recognition Elements.” *Angewandte Chemie-International Edition in English* 46 (14): 2470–2477.

doi:10.1002/anie.200603207.

Monnard, Pierre-Alain, Anastassia Kanavarioti, and D W Deamer. 2003. “Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases.” *Journal of the American Chemical Society* 125 (45) (November 12): 13734–13740. doi:10.1021/ja036465h.

Monnard, Pierre-Alain, Andrej Luptak, and D W Deamer. 2007. “Models of Primitive Cellular Life: Polymerases and Templates in Liposomes.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 362 (1486) (October 29): 1741–1750. doi:10.1126/science.276.5312.614.

Monnard, Pierre-Alain, Charles L Apel, Anastassia Kanavarioti, and D W Deamer. 2002. “Influence of Ionic Inorganic Solutes on Self-Assembly and Polymerization Processes Related to Early Forms of Life: Implications for a Prebiotic Aqueous Medium.” *Astrobiology* 2 (2): 139–152. doi:10.1089/15311070260192237.

Murtas, Giovanni. 2010. “Internal Lipid Synthesis and Vesicle Growth as a Step Toward Self-Reproduction of the Minimal Cell.” *Systems and Synthetic Biology* 4 (2) (June): 85–93. doi:10.1007/s11693-009-9048-1.

Murtas, Giovanni, Yutetsu Kuruma, Paolo Bianchini, Alberto Diaspro, and Pier Luigi Luisi. 2007. “Protein Synthesis in Liposomes with a Minimal Set of Enzymes.” *Biochemical and Biophysical Research Communications* 363 (1) (November 9): 12–17. doi:10.1016/j.bbrc.2007.07.201.

Müller, M, O Zschörnig, S Ohki, and K Arnold. 2003. “Fusion, Leakage and Surface Hydrophobicity of Vesicles Containing Phosphoinositides: Influence of Steric and Electrostatic Effects.” *The Journal of Membrane Biology* 192 (1) (March 1): 33–43. doi:10.1007/s00232-002-1062-0.

Naraoka, H, A Shimoyama, and K Harada. 1999. “Molecular Distribution of Monocarboxylic Acids in Asuka Carbonaceous Chondrites From Antarctica.” *Origins of Life and Evolution of the Biosphere* 29 (2) (March): 187–201.

Naylor, R, and P T Gilham. 1966. “Studies on Some Interactions and Reactions of Oligonucleotides in Aqueous Solution.” *Biochemistry* 5 (8) (August): 2722–2728.

Nielsen, P E. 1999. “Peptide Nucleic Acid. a Molecule with Two Identities.” *Accounts of Chemical Research* 32 (7): 624–630.

Nielsen, P E, M Egholm, R H Berg, and O Buchardt. 1991. “Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide.” *Science* 254 (5037): 1497–1500.

Noireaux, Vincent, and Albert Libchaber. 2004. “A Vesicle Bioreactor as a Step Toward an Artificial Cell Assembly.” *Proceedings of the National Academy of Sciences USA* 101 (51)

(December 21): 17669–17674. doi:10.1073/pnas.0408236101.

Nomura, Shin-ichiro M, Kanta Tsumoto, Tsutomu Hamada, Kazunari Akiyoshi, Yoichi Nakatani, and Kenichi Yoshikawa. 2003. "Gene Expression Within Cell-Sized Lipid Vesicles." *Chembiochem* 4 (11) (November 7): 1172–1175. doi:10.1002/cbic.200300630.

Nourian, Zohreh, Wouter Roelofsen, and Christophe Danelon. 2012. "Triggered Gene Expression in Fed-Vesicle Microreactors with a Multifunctional Membrane." *Angewandte Chemie-International Edition in English* 51 (13) (March 26): 3114–3118. doi:10.1002/anie.201107123.

Oberholzer, T, K H Nierhaus, and P L Luisi. 1999. "Protein Expression in Liposomes." *Biochemical and Biophysical Research Communications* 261 (2) (August 2): 238–241. doi:10.1006/bbrc.1999.0404.

Oberholzer, T, M Albrizio, and P L Luisi. 1995. "Polymerase Chain Reaction in Liposomes." *Chemistry & Biology* 2 (10) (October): 677–682.

Oberholzer, T, R Wick, P L Luisi, and C K Biebricher. 1995. "Enzymatic RNA Replication in Self-Reproducing Vesicles: an Approach to a Minimal Cell." *Biochemical and Biophysical Research Communications* 207 (1) (February 6): 250–257. doi:10.1006/bbrc.1995.1180.

Orgel, L E. 1968. "Evolution of the Genetic Apparatus." *Journal of Molecular Biology* 38 (3): 381–393.

Orgel, L E. 2004. "Prebiotic Chemistry and the Origin of the RNA World." *Critical Reviews in Biochemistry and Molecular Biology* 39 (2): 99–123. doi:10.1080/10409230490460765.

Paul, Natasha, and Gerald F Joyce. 2002. "A Self-Replicating Ligase Ribozyme." *Proceedings of the National Academy of Sciences USA* 99 (20) (October 1): 12733–12740. doi:10.1073/pnas.202471099.

Paula, S, A G Volkov, A N Van Hoek, T H Haines, and D W Deamer. 1996. "Permeation of Protons, Potassium Ions, and Small Polar Molecules Through Phospholipid Bilayers as a Function of Membrane Thickness." *Biophysical Journal* 70 (1) (January): 339–348. doi:10.1016/S0006-3495(96)79575-9.

Percec, Virgil, Daniela A Wilson, Pawaret Leowanawat, Christopher J Wilson, Andrew D Hughes, Mark S Kaucher, Daniel A Hammer, et al. 2010. "Self-Assembly of Janus Dendrimers Into Uniform Dendrimersomes and Other Complex Architectures." *Science* 328 (5981) (May 21): 1009–1014. doi:10.1126/science.1185547.

Pietrini, Adriana V, and Pier Luigi Luisi. 2004. "Cell-Free Protein Synthesis Through Solubilisate Exchange in Water/Oil Emulsion Compartments." *Chembiochem* 5 (8) (August 6): 1055–1062. doi:10.1002/cbic.200400014.

Pitsch, S, R Krishnamurthy, M Bolli, S Wendeborn, A Holzner, M Minton, C Lesueur, I Schlonvogt, B Juan, and A Eschenmoser. 1995. "Pyransoyl-RNA ('P-RNA'): Base-Pairing Selectivity and Potential to Replicate." *Helvetica Chimica Acta* 78 (7): 1621–1635.

Pohorille, Andrew, and D W Deamer. 2002. "Artificial Cells: Prospects for Biotechnology." *Trends in Biotechnology* 20 (3) (March): 123–128.

Prabahar, K J, and J P Ferris. 1997. "Adenine Derivatives as Phosphate-Activating Groups for the Regioselective Formation of 3′,5′-Linked Oligoadenylates on Montmorillonite: Possible Phosphate-Activating Groups for the Prebiotic Synthesis of RNA." *Journal of the American Chemical Society* 119 (19) (May 14): 4330–4337.

Prakash, TP, C Roberts, and C Switzer. 1997. "Activity of 2′,5′-Linked RNA in the Template-Directed Oligomerization of Mononucleotides." *Angewandte Chemie-International Edition in English* 36 (13-14): 1522–1523.

Qian, L, and E Winfree. 2011. "Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades." *Science* 332 (6034) (June 2): 1196–1201. doi:10.1126/science.1200520.

Rabinovitch-Deere, Christine A, John W K Oliver, Gabriel M Rodriguez, and Shota Atsumi. 2013. "Synthetic Biology and Metabolic Engineering Approaches to Produce Biofuels." *Chemical Reviews* 113 (7) (July 10): 4611–4632. doi:10.1021/cr300361t.

Rajamani, Sudha, Justin K Ichida, Tibor Antal, Douglas A Treco, Kevin Leu, Martin A Nowak, J W Szostak, and Irene A Chen. 2010. "Effect of Stalling After Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication." *Journal of the American Chemical Society* 132 (16) (April 28): 5880–5885. doi:10.1021/ja100780p.

Renz, M, R Lohrmann, and L E Orgel. 1971. "Catalysts for the Polymerization of Adenosine Cyclic 2′,3′-Phosphate on a Poly(U) Template." *Biochimica Et Biophysica Acta* 240 (4): 463–471.

Ricardo, Alonso, and J W Szostak. 2009. "Origin of Life on Earth." *Scientific American* 301 (3) (September): 54–61.

Rich, A. 1962. "On the Problems of Evolution and Biochemical Information Transfer." In *Horizons in Biochemistry: Albert Szent-Gyorgyi Dedicatory Volume*, edited by M Kasha and B Pullman. Academic Press.

Riley, Craig A, and Niles Lehman. 2003. "Generalized RNA-Directed Recombination of RNA." *Chemistry & Biology* 10 (12) (December): 1233–1243.

Ro, DK, EM Paradise, M Ouellet, KJ Fisher, KL Newman, JM Ndungu, KA Ho, et al. 2006. "Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast." *Nature* 440 (7086): 940–943. doi:10.1038/nature04640.

Robbins, Gregory P, Masaya Jimbo, Joe Swift, Michael J Therien, Daniel A Hammer, and Ivan J Dmochowski. 2009. "Photoinitiated Destruction of Composite Porphyrin-Protein Polymersomes." *Journal of the American Chemical Society* 131 (11) (March 25): 3872–3874. doi:10.1021/ja808586q.

Rogers, J, and G F Joyce. 2001. "The Effect of Cytidine on the Structure and Function of an RNA Ligase Ribozyme." *Rna* 7 (3) (March): 395–404.

Rohatgi, R, D P Bartel, and J W Szostak. 1996a. "Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation." *Journal of the American Chemical Society* 118 (14): 3332–3339.

Rohatgi, R, D P Bartel, and J W Szostak. 1996b. "Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3′-5′ Phosphodiester Bonds." *Journal of the American Chemical Society* 118 (14): 3340–3344.

Röthlingshöfer, Manuel, Eric Kervio, Tanja Lommel, Ulrich Plutowski, Annette Hochgesand, and Clemens Richert. 2008. "Chemical Primer Extension in Seconds." *Angewandte Chemie-International Edition in English* 47 (32) (July 28): 6065–6068. doi:10.1002/anie.v47:32.

Rushdi, A I, and B R Simoneit. 2001. "Lipid Formation by Aqueous Fischer-Tropsch-Type Synthesis Over a Temperature Range of 100 to 400 Degrees C." *Origins of Life and Evolution of the Biosphere* 31 (1-2) (February): 103–118.

Sacerdote, MG, and J W Szostak. 2005. "Semipermeable Lipid Bilayers Exhibit Diastereoselectivity Favoring Ribose." *Proceedings of the National Academy of Sciences USA* 102 (17): 6004–6008. doi:10.1073/pnas.0408440102.

Schmidt, J G, L Christensen, P E Nielsen, and L E Orgel. 1997. "Information Transfer From DNA to Peptide Nucleic Acids by Template-Directed Syntheses." *Nucleic Acids Research* 25 (23) (December 1): 4792–4796.

Schmidt, J G, P E Nielsen, and L E Orgel. 1997. "Information Transfer From Peptide Nucleic Acids to RNA by Template-Directed Syntheses." *Nucleic Acids Research* 25 (23) (December 1): 4797–4802.

Schoning, KU, P Scholz, S Guntha, X Wu, R Krishnamurthy, and A Eschenmoser. 2000. "Chemical Etiology of Nucleic Acid Structure: the Alpha-Threofuranosyl-(3′->2′) Oligonucleotide System." *Science* 290 (5495): 1347–1351.

Schramm, G, H Grotsch, and D W Pollman. 1962. "Non-Enzymatic Synthesis of Polysaccharides, Nucleosides and Nucleic Acids and the Origin of Self-Reproducing Systems." *Angewandte Chemie-International Edition in English* 1 (1) (September 19): 1–64.

Schrum, Jason P, Alonso Ricardo, Mathangi Krishnamurthy, J C Blain, and J W Szostak. 2009.

"Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2′-Amino-2′,3′-Dideoxyribonucleoside-5′-Phosphorimidazolide Monomers." *Journal of the American Chemical Society* 131 (40) (October 14): 14560–14570. doi:10.1021/ja906557v.

Sczepanski, Jonathan T, and Gerald F Joyce. 2012. "Synthetic Evolving Systems That Implement a User-Specified Genetic Code of Arbitrary Design." *Chemistry & Biology* 19 (10) (October 26): 1324–1332. doi:10.1016/j.chembiol.2012.08.017.

Shechner, David M, and D P Bartel. 2011. "The Structural Basis of RNA-Catalyzed RNA Polymerization." *Nature Structural & Molecular Biology* 18 (9) (September): 1036–1042. doi:10.1038/nsmb.2107.

Shechner, David M, Robert A Grant, Sarah C Bagby, Yelena Koldobskaya, Joseph A Piccirilli, and D P Bartel. 2009. "Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme." *Science* 326 (5957) (November 27): 1271–1275. doi:10.1126/science.1174676.

Shimizu, Y, A Inoue, Y Tomari, T Suzuki, T Yokogawa, K Nishikawa, and T Ueda. 2001. "Cell-Free Translation Reconstituted with Purified Components." *Nature Biotechnology* 19 (8) (August): 751–755. doi:10.1038/90802.

Sievers, D, and G von Kiedrowski. 1994. "Self-Replication of Complementary Nucleotide-Based Oligomers." *Nature* 369 (6477) (May 19): 221–224. doi:10.1038/369221a0.

Sintim, HO, and ET Kool. 2006. "Enhanced Base Pairing and Replication Efficiency of Thiothymidines, Expanded-Size Variants of Thymidine." *Journal of the American Chemical Society* 128 (2): 396–397. doi:10.1021/ja0562447.

Staneva, Galya, Miglena I Angelova, and Kamen Koumanov. 2004. "Phospholipase A2 Promotes Raft Budding and Fission From Giant Liposomes." *Chemistry and Physics of Lipids* 129 (1) (April): 53–62. doi:10.1016/j.chemphyslip.2003.11.005.

Strulson, Christopher A, Rosalynn C Molden, Christine D Keating, and Philip C Bevilacqua. 2012. "RNA Catalysis Through Compartmentalization." *Nature Chemistry* 4 (11) (November): 941–946. doi:10.1038/nchem.1466.

Stuetz, Jan A Rojas, Eric Kervio, Christopher Deck, and Clemens Richert. 2007. "Chemical Primer Extension: Individual Steps of Spontaneous Replication." *Chemistry & Biodiversity* 4 (4) (April): 784–802. doi:10.1002/cbdv.200790064.

Sulston, J, R Lohrmann, L E Orgel, and H T Miles. 1968. "Nonenzymatic Synthesis of Oligoadenylates on a Polyuridylic Acid Template." *Proceedings of the National Academy of Sciences USA* 59 (3) (March): 726–733.

Sunami, Takeshi, Filippo Caschera, Yuuki Morita, Taro Toyota, Kazuya Nishimura, Tomoaki Matsuura, Hiroaki Suzuki, Martin M Hanczyc, and Tetsuya Yomo. 2010. "Detection of Association and Fusion of Giant Vesicles Using a Fluorescence-Activated Cell Sorter."

Langmuir 26 (19) (October 5): 15098–15103. doi:10.1021/la102689v.

Sunami, Takeshi, Kazufumi Hosoda, Hiroaki Suzuki, Tomoaki Matsuura, and Tetsuya Yomo. 2010. "Cellular Compartment Model for Exploring the Effect of the Lipidic Membrane on the Kinetics of Encapsulated Biochemical Reactions." *Langmuir* 26 (11) (June 1): 8544–8551. doi:10.1021/la904569m.

Szathmary, E, and J M Smith. 1995. "The Major Evolutionary Transitions." *Nature* 374 (6519) (March 16): 227–232. doi:10.1038/374227a0.

Szostak, J W. 2011. "An Optimal Degree of Physical and Chemical Heterogeneity for the Origin of Life?." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 366 (1580) (October 27): 2894–2901. doi:10.1098/rstb.2011.0140.

Szostak, J W. 2012. "The Eightfold Path to Non-Enzymatic RNA Replication." *Journal of Systems Chemistry* 3 (September 7): 2.

Szostak, J W, D P Bartel, and P L Luisi. 2001. "Synthesizing Life." *Nature* 409 (6818): 387–390.

Takahashi, Hiroshi, Yoshiyuki Kageyama, Kensuke Kurihara, Katsuto Takakura, Shigeru Murata, and Tadashi Sugawara. 2009. "Autocatalytic Membrane-Amplification on a Pre-Existing Vesicular Surface." *Chemical Communications* 46 (46): 8791. doi:10.1039/c0cc02758h.

Takakura, Katsuto, and Tadashi Sugawara. 2004. "Membrane Dynamics of a Myelin-Like Giant Multilamellar Vesicle Applicable to a Self-Reproducing System." *Langmuir* 20 (10) (May 11): 3832–3834.

Takakura, Katsuto, Taro Toyota, and Tadashi Sugawara. 2003. "A Novel System of Self-Reproducing Giant Vesicles." *Journal of the American Chemical Society* 125 (27) (July 9): 8134–8140. doi:10.1021/ja029379a.

Tanford, Charles. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. John Wiley & Sons.

Tawfik, D S, and A D Griffiths. 1998. "Man-Made Cell-Like Compartments for Molecular Evolution." *Nature Biotechnology* 16 (7) (July): 652–656. doi:10.1038/nbt0798-652.

Terasawa, Hidetoshi, Kazuya Nishimura, Hiroaki Suzuki, Tomoaki Matsuura, and Tetsuya Yomo. 2012. "Coupling of the Fusion and Budding of Giant Phospholipid Vesicles Containing Macromolecules." *Proceedings of the National Academy of Sciences USA* 109 (16) (April 17): 5942–5947. doi:10.1073/pnas.1120327109.

Testa, S M, M D Disney, D H Turner, and R Kierzek. 1999. "Thermodynamics of RNA-RNA Duplexes with 2- or 4-Thiouridines: Implications for Antisense Design and Targeting a Group I Intron." *Biochemistry* 38 (50) (December 14): 16655–16662.

Tohidi, M, W S Zielinski, C H Chen, and L E Orgel. 1987. “Oligomerization of 3′-Amino-3′-Deoxyguanosine-5′-Phosphorimidazolidate on a D(CpCpCpCpC) Template.” *Journal of Molecular Evolution* 25: 97–99.

Treyer, M, P Walde, and T Oberholzer. 2002. “Permeability Enhancement of Lipid Vesicles to Nucleotides by Use of Sodium Cholate: Basic Studies and Application to an Enzyme-Catalyzed Reaction Occurring Inside the Vesicles.” *Langmuir* 18 (4) (February 19): 1043–1050. doi:10.1021/la011111u.

Ura, Yasuyuki, John M Beierle, Luke J Leman, L E Orgel, and M Reza Ghadiri. 2009. “Self-Assembling Sequence-Adaptive Peptide Nucleic Acids.” *Science* 325 (5936) (July 3): 73–77. doi:10.1126/science.1174577.

Urabe, Hiroya, Norikazu Ichihashi, Tomoaki Matsuura, Kazufumi Hosoda, Yasuaki Kazuta, Hiroshi Kita, and Tetsuya Yomo. 2010. “Compartmentalization in a Water-in-Oil Emulsion Repressed the Spontaneous Amplification of RNA by Q Beta Replicase.” *Biochemistry* 49 (9) (March 9): 1809–1813. doi:10.1021/bi901805u.

Vaidya, Nilesh, Michael L Manapat, Irene A Chen, Ramon Xulvi-Brunet, Eric J Hayden, and Niles Lehman. 2012. “Spontaneous Network Formation Among Cooperative RNA Replicators.” *Nature* 491 (7422) (November 1): 72–77. doi:10.1038/nature11549.

Vicens, Quentin, and Thomas R Cech. 2009. “A Natural Ribozyme with 3′,5′ RNA Ligase Activity.” *Nature Chemical Biology* 5 (2) (February 4): 97–99. doi:10.1038/nchembio.136.

Vogel, Stephanie R, Christopher Deck, and Clemens Richert. 2005. “Accelerating Chemical Replication Steps of RNA Involving Activated Ribonucleotides and Downstream-Binding Elements.” *Chemical Communications* (39) (October 21): 4922–4924. doi:10.1039/b510775j.

Walde, P, A Goto, P Monnard, M Wessicken, and P L Luisi. 1994. “Oparin’s Reactions Revisited: Enzymatic Synthesis if Poly(Adenylic Acid) in Micelles and Self-Reproducing Vesicles.” *Journal of the American Chemical Society* 116 (May 20): 7541–7547.

Wang, Qing S, Leslie K L Cheng, and Peter J Unrau. 2011. “Characterization of the B6.61 Polymerase Ribozyme Accessory Domain.” *Rna* 17 (3) (March): 469–477. doi:10.1261/rna.2495011.

Wasner, M, D Arion, G Borkow, A Noronha, AH Uddin, MA Parniak, and MJ Damha. 1998. “Physicochemical and Biochemical Properties of 2′,5′-Linked RNA and 2′,5′-RNA:3′,5′-RNA ‘Hybrid’ Duplexes.” *Biochemistry* 37 (20): 7478–7486.

Weimann, B J, R Lohrmann, L E Orgel, H Schneider-Bernloehr, and J E Sulston. 1968. “Template-Directed Synthesis with Adenosine-5'-Phosphorimidazolide.” *Science* 161 (3839) (July 26): 387.

Wittung, P, P E Nielsen, O Buchardt, M Egholm, and B Nordén. 1994. "DNA-Like Double Helix Formed by Peptide Nucleic Acid." *Nature* 368 (6471) (April 7): 561–563. doi:10.1038/368561a0.

Wochner, Aniela, James Attwater, Alan Coulson, and Philipp Holliger. 2011. "Ribozyme-Catalyzed Transcription of an Active Ribozyme." *Science* 332 (6026) (April 8): 209–212. doi:10.1126/science.1200752.

Woese, Carl R. 1967. *The Genetic Code: the Molecular Basis for Genetic Expression*. Harper & Row.

Wojciechowski, Filip, and Christian J Leumann. 2011. "Alternative DNA Base-Pairs: From Efforts to Expand the Genetic Code to Potential Material Applications." *Chemical Society Reviews* 40 (12): 5669. doi:10.1039/c1cs15027h.

Wright, MC, and GF Joyce. 1997. "Continuous in Vitro Evolution of Catalytic Function." *Science* 276 (5312): 614–617.

Wu, T F, and L E Orgel. 1992a. "Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 2. Templates Containing Cytidine and Guanosine Residues." *Journal of the American Chemical Society* 114 (14): 5496–5501.

Wu, T F, and L E Orgel. 1992b. "Nonenzymatic Template-Directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides." *Journal of the American Chemical Society* 114 (1): 317–322.

Wu, T F, and L E Orgel. 1992c. "Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 3. Incorporation of Adenosine and Uridine Residues." *Journal of the American Chemical Society* 114 (21) (October 7): 7963–7969.

Xia, TB, J SantaLucia, ME Burkard, R Kierzek, SJ Schroeder, XQ Jiao, C Cox, and DH Turner. 1998. "Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs." *Biochemistry* 37 (42): 14719–14735.

Yoshimoto, Makoto, Shaoqing Wang, Kimitoshi Fukunaga, Didier Fournier, Peter Walde, Ryoichi Kuboi, and Katsumi Nakao. 2005. "Novel Immobilized Liposomal Glucose Oxidase System Using the Channel Protein OmpF and Catalase." *Biotechnology and Bioengineering* 90 (2) (April 20): 231–238. doi:10.1002/bit.20422.

Yu, Hanyang, Su Zhang, and John C Chaput. 2012. "Darwinian Evolution of an Alternative Genetic System Provides Support for TNA as an RNA Progenitor." *Nature Chemistry* 4 (3) (March): 183–187. doi:10.1038/nchem.1241.

Yu, W, K Sato, M Wakabayashi, T Nakaishi, E P Ko-Mitamura, Y Shima, I Urabe, and T Yomo. 2001. "Synthesis of Functional Protein in Liposome." *Journal of Bioscience and Bioengineering* 92 (6): 590–593.

Yuen, G U, and K A Kvenvolden. 1973. “Monocarboxylic Acids in Murray and Murchison Carbonaceous Meteorites.” *Nature* 246 (5431) (November 7): 301–303.

Zaher, Hani S, and Peter J Unrau. 2007. “Selection of an Improved RNA Polymerase Ribozyme with Superior Extension and Fidelity.” *Rna* 13 (7) (July): 1017–1026. doi:10.1261/rna.548807.

Zhang, S, J C Blain, D Zielinska, S Gryaznov, and J W Szostak. “Fast and Accurate Non-Enzymatic Copying of an RNA-Like Synthetic Genetic Polymer.” *Submitted*.

Zhang, S, N Zhang, J C Blain, and J W Szostak. 2013. “Synthesis of N3′-P5′-Linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension.” *Journal of the American Chemical Society* 135 (January 7): 924–932. doi:10.1021/ja311164j.

Zhu, Ting F, and J W Szostak. 2009. “Coupled Growth and Division of Model Protocell Membranes.” *Journal of the American Chemical Society* 131 (15) (April 22): 5705–5713. doi:10.1021/ja900919c.

Zhu, Ting F, Katarzyna Adamala, Na Zhang, and J W Szostak. 2012. “Photochemically Driven Redox Chemistry Induces Protocell Membrane Pearling and Division.” *Proceedings of the National Academy of Sciences USA* 109 (25) (June 19): 9828–9832. doi:10.1073/pnas.1203212109.

Zielinski, W S, and L E Orgel. 1985. “Oligomerization of Activated Derivatives of 3′-Amino-3′-Deoxyguanosine on Poly(C) and Poly(dC) Templates.” *Nucleic Acids Research* 13 (7) (April 11): 2469–2484.

Zielinski, W S, and L E Orgel. 1987. “Oligoaminonucleoside Phosphoramidates. Oligomerization of Dimers of 3′-Amino-3′-Deoxy-Nucleotides (GC and CG) in Aqueous Solution.” *Nucleic Acids Research* 15 (4): 1699–1715.

Chapter Two: The chemical mechanism of non-enzymatic RNA polymerization

J. Craig Blain, Isabel M. Vogt and Jack W. Szostak

Abstract

The non-enzymatic self-replication of RNA may have played a critical role in the early evolution of life and it could be used as a method of genome replication in artificial protocells. However, the chemical mechanism of this reaction is not well understood. Using the template-directed addition of guanosine 5′-phosphor-2-methylimidazolide (2-MeImpG) to an RNA primer as a model reaction, we first interrogated the mechanism of Mg^{2+} catalysis. By comparing Mg^{2+} to $Co(NH_3)_6^{3+}$ we found that it likely plays two roles: stabilization of 2-MeImpG binding through outer shell or diffuse charge interactions, and enhancement of phosphoryl transfer through inner shell contacts. From the absence of a solvent deuterium kinetic isotope effect we were able to rule out general base catalysis by a Mg^{2+}-bound hydroxide ion. The pH dependence of the reaction fit best to a model with two ionizable groups, one with a pK_a of 6.9±0.1 requiring deprotonation and one with a pK_a of 8.2±0.1 requiring protonation, resulting in a maximum rate at pH 7.5. Using a 2′-deoxy terminated primer and using Ca^{2+} instead of Mg^{2+} did not have a large effect on either pK_a, indicating that the 3′-hydroxyl is likely protonated during attack. Therefore, we favor a model where Mg^{2+} coordinates the 2-MeImpG phosphate and primer 3′-hydroxyl to orient them for attack and in-line displacement of a protonated 2-methylimidazole leaving group. This model will help guide future efforts to optimize the reaction and incorporate it into fatty acid vesicles.

Introduction

According to the RNA world hypothesis (Gilbert 1986), modern DNA, RNA and protein based cells were predated by simpler life forms that used RNA to both store genetic information and catalyze biochemical reactions. To support this hypothesis, ribozymes have been isolated through *in vitro* selection that are able to catalyze RNA-dependent RNA polymerization (Johnston et al. 2001; Wochner et al. 2011), a critical reaction for self-replication. However, these ribozymes are close to 200 nucleotides long and a single polymerase molecule cannot make a copy of itself. Therefore, for life to have started with ribozyme catalyzed self-replication two of these complex polymerase ribozymes would have had to form in close proximity through random nucleotide polymerization on the early earth. Such an event would have been so improbable that the RNA world itself likely evolved from an even simpler system. Pioneering work by Leslie Orgel and colleagues has shown that highly activated nucleotides can copy RNA templates without any enzymatic catalysis (Inoue and Orgel 1983). This process could have bridged the gap between prebiotic chemistry and more efficient ribozyme catalyzed replication.

Specifically, Orgel discovered that ribonucleotides with a 2-methylimidazole activated phosphate, known as phosphor-2-methylimidazolides, will polymerize into 3′-5′-linked oligomers when incubated with a complementary template (Inoue and Orgel 1983). This reaction is kinetically complex, involving many binding equilibria and ligation rates (Kanavarioti et al. 1993). Therefore, a simpler system was developed wherein the activated nucleotides extend from a labeled primer bound to a template and the reaction is followed be gel electrophoresis (Wu and Orgel 1992a). Since it allows for the clear kinetic dissection

of an individual nucleotide ligation reaction, this system has become the standard model for studying non-enzymatic, template-directed polymerization.

Since the initial experiments performed in the late 1960s (Sulston et al. 1968), much has been learned about non-enzymatic RNA polymerization. The reaction requires Mg^{2+} as a catalyst, although other divalent cations including Zn^{2+} and Pb^{2+} can be used (Lohrmann, Bridson, and Orgel 1980). Subtle changes to the activating group can have a large effect on the reaction. Other imidazole derivatives, even 2-ethylimidazole, do not give rates as fast as 2-methylimidazole (Inoue and Orgel 1981) and very few other activating groups have been used successfully (Prabahar and Ferris 1997; Deck, Jauker, and Richert 2011). Interestingly, the polymerization of imidazole activated nucleotides catalyzed by Mg^{2+} or Pb^{2+} gives primarily 2′-5′-linked products, whereas 3′-5′-linked products are obtained from imidazole activation and Zn^{2+} catalysis (Lohrmann, Bridson, and Orgel 1980) or 2-methylimidazole activation and Mg^{2+} catalysis (Inoue and Orgel 1981). Furthermore, different base pairs give different rates of reaction. Guanosine and cytidine monomers polymerize the fastest, presumably since they form a stronger base pair and guanosine is the faster of the two since it has stronger base stacking (Wu and Orgel 1992b). Adenosine and uridine monomers polymerize more slowly since they form a weaker base pair (Wu and Orgel 1992c).

Despite decades of work, non-enzymatic replication of RNA sequences has still not been achieved. However, progress has been hindered by the *ad hoc* approach taken to improving the efficiency of the reaction. Although the kinetics of non-enzymatic polymerization have been studied extensively (Kanavarioti et al. 1993), the basic chemical mechanism of ligation is not well understood (Figure 2.1). In particular, to rationally

optimize the reaction rate it is critical to address the following questions: (i) What is the role of Mg^{2+} in the reaction? For example, it could activate the 3′-hydroxyl group, orient the 3′-hydroxyl nucleophile and phosphate electrophile, stabilize negative charge build up on the phosphorane transition state, provide a bound hydroxide as a general base, or improve monomer binding affinity. (ii) Is the phosphoryl transfer reaction dissociative, associative or concerted? If it is concerted, is the transition state loose or tight? (iii) What ionization state is required for the groups involved in the reaction, in particular the 3′-hydroxyl nucleophile and 2-methylimidazole leaving group?

By performing some basic physical organic chemistry experiments we were able to address some of these questions. The results have helped us understand the chemical context required for non-enzymatic RNA polymerization in the origin of life and will help guide efforts to develop a self-replicating RNA system. They also suggest ways to optimize the reaction and reduce the dependence on Mg^{2+} so that it can be incorporated into a fatty acid vesicle encapsulated protocell system (see Chapter 1). Furthermore, non-enzymatic models of biological reactions have proven to be important comparisons to better understand the role that enzymes play (Loeb and Kunkel 1982; Lassila, Zalatan, and Herschlag 2011).

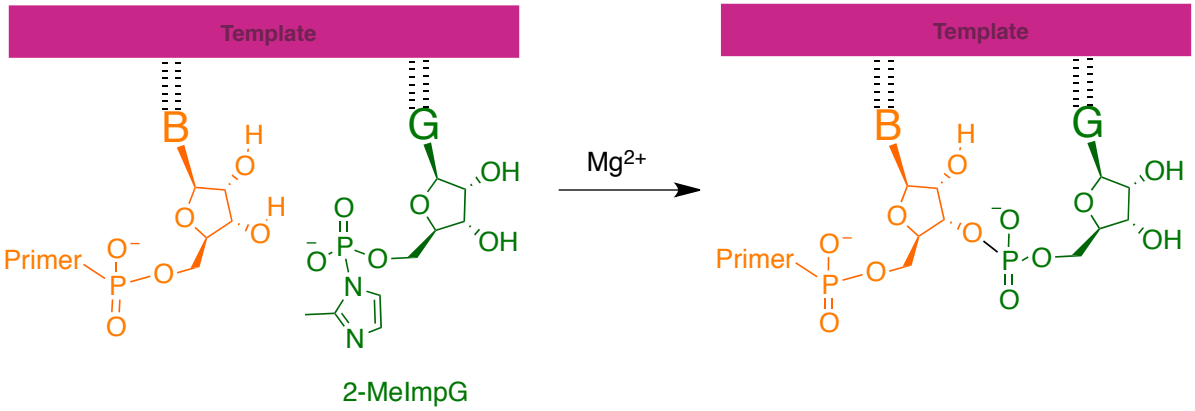


Figure 2.1. Scheme for a primer extension reaction with an RNA primer and guanosine 5′-phosphor-2-methylimidazolide (2-MeImpG) monomer.

Results

A model reaction for non-enzymatic RNA polymerization

In order to study the mechanism of a reaction it is important to have a simple model system that allows for direct measurement of the reaction kinetics. For the non-enzymatic polymerization of ribonucleotides we chose to study the 5′→3′ addition of guanosine 5′-phosphor-2-methylimidazolide (2-MeImpG) to an RNA primer on an RNA C_4 template (Figure 2.2A). The template had an additional two adenosine residues on its 5′-terminus to assist with stacking of the template residues. The primer was labeled with a fluorophore and the reaction products were analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 2.2B). At each time point the bands for the primer and extension products were integrated. To determine the pseudo-first order rate constant (k_{obs}) for the reaction the fraction of the total integration represented by primer was plotted against time and fit to a single exponential decay (Figure 2.2C).

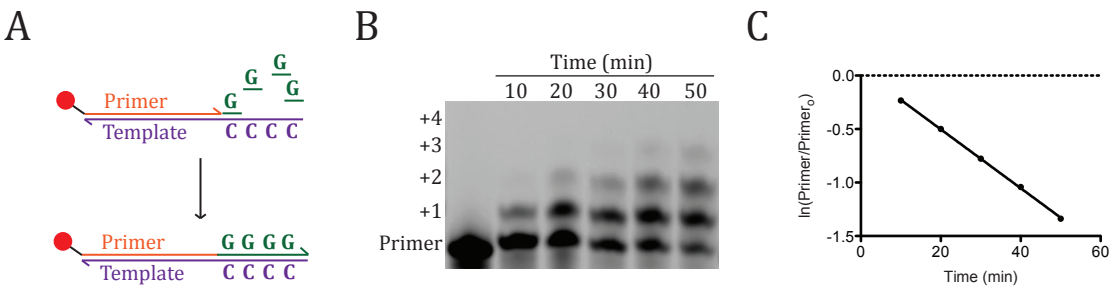


Figure 2.2. Non-enzymatic primer extension reaction assay. (A) Schematic diagram of the non-enzymatic primer extension reaction of 2-MeImpG on an RNA C_4 template. (B) Example PAGE analysis of a primer extension reaction. (C) Example plot of the natural logarithm of the fraction of unreacted primer remaining versus time. The pseudo-first order rate constant (k_{obs}) was taken as the negative slope of a linear fit to this plot.

Magnesium and monomer binding

First, we measured the dependence of the reaction rate on the concentration of $MgCl_2$ at 50 mM 2-MeImpG. The dependence fit well to a single binding site model with a K_D of 150±20 mM (estimate ± standard error) and a maximum rate constant of 6.1±0.4 h^{-1} (R^2 > 0.99), although we were not able to reach a saturating concentration of $MgCl_2$ (Figure 2.3A). In the absence of added Mg^{2+} the rate constant was 0.02 h^{-1}. For subsequent experiments we wanted to limit the concentration of $MgCl_2$ while maintaining a rapid extension rate so we decided on a concentration of 50 mM, which gave a rate constant of 1.7 h^{-1} and a half-time of 24 min. We next measured the dependence of the reaction rate on the concentration of 2-MeImpG. Unexpectedly, the reaction rate fit well to a single binding site model up to 50 mM and then decreased steadily (Figure 2.3B), possibly due to a competition for Mg^{2+} binding between bound and free monomer. Therefore, we used 50 mM 2-MeImpG in all subsequent reactions.

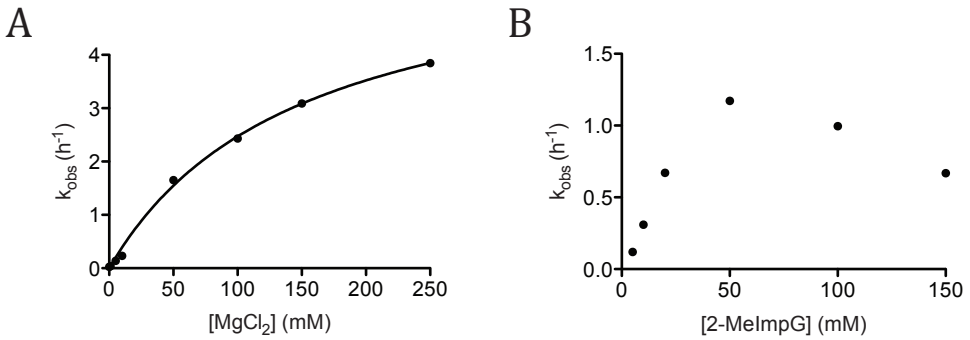


Figure 2.3. Dependence of the rate of primer extension on the concentration of $MgCl_2$ and 2-MeImpG. (A) The pseudo-first order rate constant (k_{obs}) was measured with 50 mM 2-MeImpG and varying concentrations of $MgCl_2$. The plot was fit to a single site binding model with a K_D of 150±20 mM (estimate ± standard error) and k_{max} of 6.1±0.4 h^{-1} (R^2 > 0.99). (B) The rate constant was measured with 50 mM $MgCl_2$ and varying concentrations of 2-MeImpG. All reactions were performed at pH 7.5.

Inner versus outer shell contacts in Mg^{2+} catalysis

In aqueous solution Mg^{2+} ions have an inner hydration shell composed of six water molecules bound in an octahedral geometry followed by a less ordered outer shell (Bowman et al. 2012). Three types of interactions are possible between Mg^{2+} and RNA: (i) it can partially dehydrate and make direct inner shell contacts with the RNA, typically to phosphate groups; (ii) it can contact the RNA in its outer hydration shell, or; (iii) it can form diffuse charge interactions outside of its hydration shells (Draper, Grilley, and Soto 2005). To differentiate between inner and outer shell contacts we compared the catalytic abilities of Mg^{2+} and cobalt(III) hexamine. Similar to Mg^{2+} hydration, $Co(NH_3)_6^{3+}$ has an octahedral inner shell of six ammonia molecules; however, they are stably bound and do not exchange

to allow other direct contacts to Co^{3+}. Therefore, $Co(NH_3)_6^{3+}$ can effectively substitute with Mg^{2+} for outer shell and diffuse ion contacts, but not inner shell contacts (Nesbitt, Hegg, and Fedor 1997). We found that $Co(NH_3)_6^{3+}$ could catalyze non-enzymatic primer extension with a k_{max} of 0.29±0.02 h^{-1} and K_D of 8±2 mM (R^2 = 0.97) (Figure 2.4A). Therefore, compared to Mg^{2+}, $Co(NH_3)_6^{3+}$ saturated at a lower concentration, but also at a lower maximum rate.

The limited catalysis by $Co(NH_3)_6^{3+}$ could be explained in at least two ways: (i) $Co(NH_3)_6^{3+}$ could substitute for Mg^{2+}, but was not as effective; or (ii) multiple Mg^{2+} ions used either inner shell or outer shell and diffuse charge interactions to catalyze the reaction and $Co(NH_3)_6^{3+}$ could only substitute for the outer shell or diffuse interactions. To differentiate between these possibilities we performed a Mg^{2+} titration experiment in the presence and absence of 50 mM $Co(NH_3)_6^{3+}$ (Horton and DeRose 2000). We found that $Co(NH_3)_6^{3+}$ enhanced the reaction rate at Mg^{2+} concentrations below 250 mM, indicating that Mg^{2+} and $Co(NH_3)_6^{3+}$ do not compete for binding to the same site (Figure 2.4B).

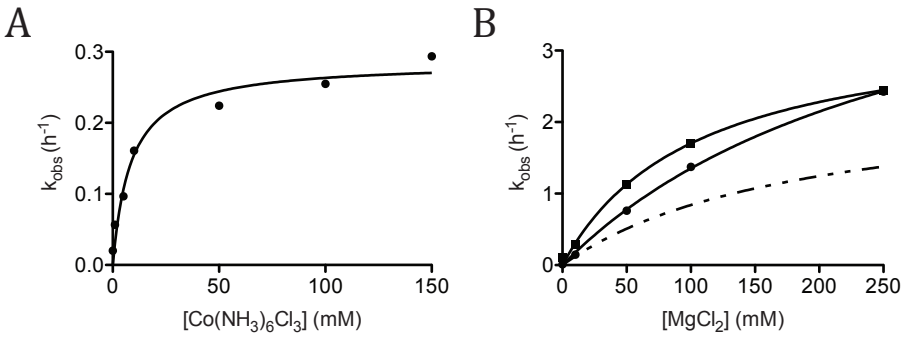


Figure 2.4. Comparison of Mg^{2+} and $Co(NH_3)_6^{3+}$ as catalysts for non-enzymatic primer extension. (A) Dependence of the reaction rate on the concentration of $Co(NH_3)_6Cl_3$ in the absence of $MgCl_2$. The plot was fit to a single site binding model with a K_D of 8±2 mM and a

Figure 2.4 (continued). k_{max} of 0.29±0.02 h^{-1} (R^2 = 0.97). (B) Dependence of the reaction rate on the concentration of $MgCl_2$ in the absence (•) and presence (■) of 50 mM $Co(NH_3)_6Cl_3$. The dashed line is the curve predicted in the presence of 50 mM $Co(NH_3)_6Cl_3$ if it was a competitive inhibitor of $MgCl_2$.

Solvent deuterium kinetic isotope effect

Phosphoryl transfer reactions catalyzed by Mg^{2+} often involve general base catalysis by a Mg^{2+} bound hydroxide ion (Yang, Lee, and Nowotny 2006). To test for this we measured the solvent deuterium kinetic isotope effect (SDKIE) for the primer extension reaction. In this experiment, the mole fraction of deuterium in the water solvent is gradually increased. If a proton transfer occurs in the rate-determining step of the reaction, exchange of the proton for deuterium will give a primary kinetic isotope effect that can slow the rate by up to a factor of six (Venkatasubban and Schowen 1984). We did not observe a decrease in rate up to roughly 90% deuterium (Figure 2.5A), demonstrating that there was no general acid-base catalysis in the rate-determining step. A similar result was observed using the imidazole-activated monomer guanosine 5′-phosphor-2-methylimidazole (ImpG) (Figure 2.5B).

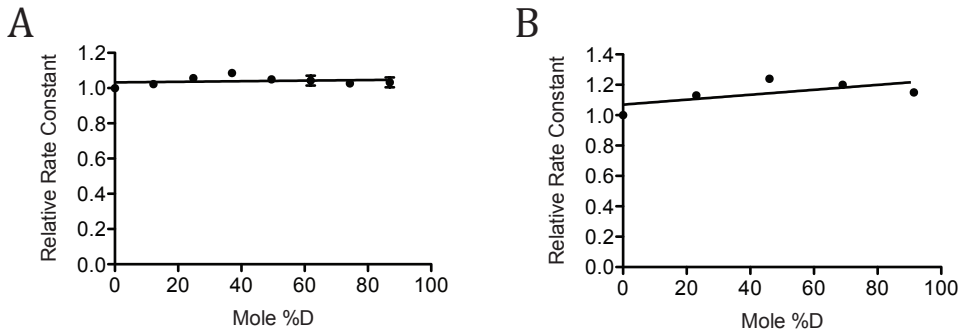


Figure 2.5. Solvent deuterium kinetic isotope effect. The rate constant for the primer extension reaction was measured at different mole fractions of deuterium in the solvent and normalized to 0% deuterium. (A) Using 2-MeImpG as the monomer. Error bars represent SEM (n = 2). (B) Using ImpG as the monomer.

pH dependence of non-enzymatic primer extension

To determine the required ionization state for each group involved in the reaction we measured the rate of the reaction from pH 5.5 to 10. The profile fit well to a model with two ionizable groups (Bevilacqua 2003), one with a pK_a of 8.2±0.1 requiring protonation and another with a pK_a of 6.9±0.1 requiring deprotonation, and a rate constant of 2.7±0.3 h^{-1} when in the optimal protonation state (Figure 2.6A). However, an alternate model with a single group requiring protonation and deprotonation at different points is also possible (Breslow 1993). To determine which groups were being ionized we changed the conditions of the reaction and measured the effect on the pH-rate profile (Table 2.1). To test if the 3′- or 2′-hydroxyl groups had to be deprotonated we used a primer with a 3′-terminal 2′-deoxy residue (Figure 2.6B). Although the overall reaction rate was lower, the fitted pK_a values were very similar. We then tested whether either of the pK_a values were affected by binding to Mg^{2+} by using Ca^{2+} instead (Figure 2.6C). Again, both pK_a values were very

similar. Finally, we tested the imidazole-activated monomer ImpG (Figure 2.6D). In this case, pK_a values of 6.6±0.1 and 9.6±0.1 were observed, with a maximal rate constant of 0.119±0.006 h^{-1}, indicating that either the higher pK_a shifted upwards or the reaction mechanism changed and it represented a different group.

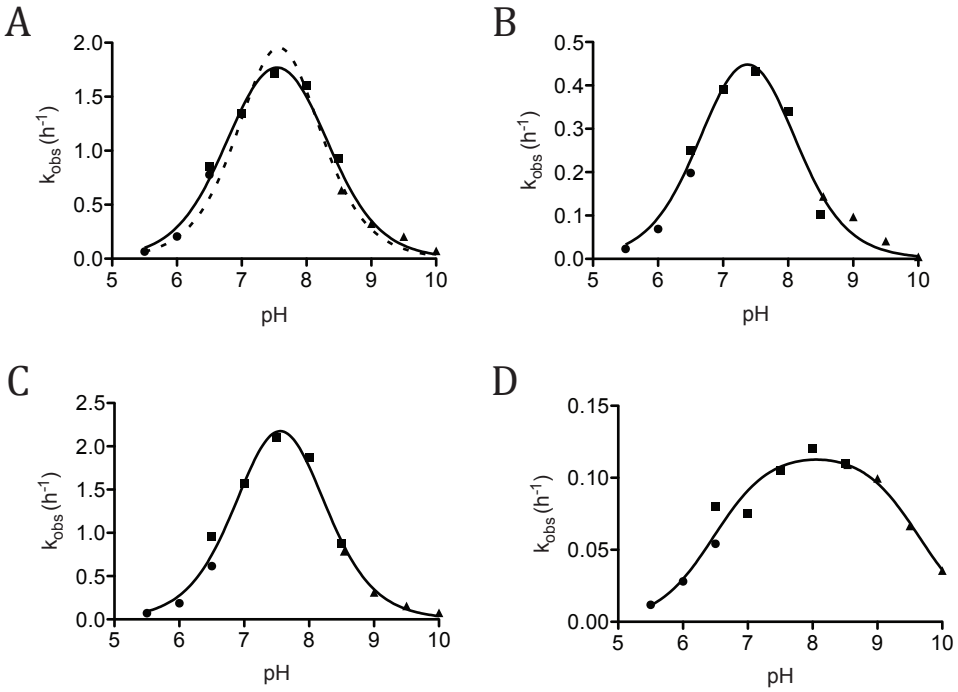


Figure 2.6. pH dependence of the primer extension reaction under different conditions. Solid lines represent the best fit to a model with two ionizable groups with different pK_a values (eq. 2.2). (A) Using the standard reaction conditions: 50 mM $MgCl_2$, 50 mM 2-MeImpG, and an RNA primer (R^2 = 0.98). The dotted line represents the best fit when the two pK_a values were forced to be equal (R^2 = 0.95). (B) Using a primer with a 3′-terminal 2′-deoxy residue (R^2 = 0.97). (C) Using $CaCl_2$ instead of $MgCl_2$ (R^2 = 0.98). (D) Using the imidazole-activated ImpG monomer instead of 2-MeImpG (R^2 = 0.95). Three different buffers were used: MES (•), HEPES (■), and CHES (▲).

Table 2.1. Best fit parameters for the pH dependence of the primer extension reaction. Values are presented as estimate ± standard error.

Catalyst	Primer 3′-End	Monomer	pK_{a1}	pK_{a2}	k (h^{-1})
$MgCl_2$	Ribose	2-MeImpG	6.9±0.1	8.2±0.1	2.7±0.3
$MgCl_2$	2′-Deoxy	2-MeImpG	6.8±0.1	7.9±0.1	0.7±0.1
$CaCl_2$	Ribose	2-MeImpG	7.2±0.1	7.9±0.1	4.2±0.8
$MgCl_2$	Ribose	ImpG	6.6±0.1	9.6±0.1	0.119±0.006

Discussion

Our results address many longstanding questions about the chemical mechanism of non-enzymatic primer extension with 2-MeImpG, a model system for RNA self-replication in the origin of life. Since the reaction is template-dependent and catalyzed by Mg^{2+}, it is generally assumed to proceed through the formation of a 2-MeImpG-Mg^{2+}-primer-template complex followed by nucleophilic attack of the primer 3′-hydroxyl on the phosphate of 2-MeImpG to displace 2-methylimidazole and form a new phosphodiester bond (Kanavarioti et al. 1989; Wu and Orgel 1992a). In support of this view, we found that the dependence of the reaction rate on the concentration of Mg^{2+} and 2-MeImpG was consistent with each binding to the primer-template prior to the phosphoryl transfer step (Figure 2.3). To the best of our knowledge the Mg^{2+} dependence of this reaction has not previously been reported, but a similar binding interaction was observed for Mg^{2+} in the non-enzymatic ligation of a triphosphate activated oligonucleotide (Rohatgi, Bartel, and Szostak 1996).

Unexpectedly, we also found that 2-MeImpG began to inhibit the reaction at concentrations above 50 mM (Figure 2.3B). This effect could have been the result of free 2-

MeImpG competing with the primer-template-2-MeImpG complex for binding to Mg^{2+}. Binding between 2-MeImpG and Mg^{2+} has been reported (Kanavarioti et al. 1989); however, the K_D was estimated at 0.9 M, which is too weak to compete significantly with the complex K_D of 150±20 mM. Alternatively, as the average occupancy of the C_4 template increased stacking interactions between 2-MeImpG monomers could have forced the complex to take on an unfavorable geometry. It has previously been reported that the reaction rate plateaus at 50 mM 2-MeImpG, but only single time points were compared for each monomer concentration so precise rates could not be measured (Wu and Orgel 1992a). Furthermore, that study used 200 mM $MgCl_2$ whereas we used 50 mM and the interdependence of Mg^{2+} and 2-MeImpG binding has not been studied yet.

We identified at least two roles for Mg^{2+} in the reaction, one that could be performed by $Co(NH_3)_6^{3+}$ and one that could not (Figure 2.4). The former is likely the stabilization of 2-MeImpG binding through outer shell or diffuse charge interactions, and the latter is likely catalysis of the reaction through inner shell contacts. Although the k_{obs} *v.* $[Mg^{2+}]$ plot fit well to a single binding site model (Figure 2.3A), it is difficult to differentiate that from multiple binding sites. Indeed, the plot was slightly sigmoidal, possibly indicating positive cooperativity between the stabilization of 2-MeImpG binding and binding to the reaction center. Since no SDKIE was observed (Figure 2.5), we were able to rule out general base catalysis by a Mg^{2+}-bound hydroxide ion in the rate-determining step. In contrast, DNA and RNA polymerases have a non-linear SDKIE, indicating multiple proton transfers in the rate-determining step (Castro et al. 2007), and the class I ligase ribozyme has a single proton transfer (Shechner and Bartel 2011).

Throughout this work we have made the assumption that the reaction proceeds through a single step with concerted attack of the 3′-hydroxyl and leaving of 2-methylimidazole to form a phosphodiester bond (Figure 2.1). We could rule out a dissociative mechanism to form a metaphosphate intermediate since it is only favorable with phosphate monoesters (Lassila, Zalatan, and Herschlag 2011). Conversely, an associative mechanism with the formation of a stable pentacoordinate intermediate is possible and would significantly alter the interpretation of the data if dissociation of the intermediate was rate limiting or if pseudorotation of the intermediate was involved. Experimentally testing for an associative mechanism is notoriously difficult and usually involves phosphotriesters, which have limited relevance to this work (Kluger and Westheimer 1969). More recently, Lönnberg *et al.* have used a breakpoint in the Brønsted plot for the activating group pK_a as evidence for a stable intermediate (Lönnberg, Strömberg, and Williams 2004) and a similar study would be helpful for our system. However, since the di-anionic phosphoranes produced from nucleophilic attack on di-substituted phosphates are generally unstable (Perreault and Anslyn 1997) and since protonated 2-methylimidazole is a good leaving group, an associative mechanism is unlikely for the non-enzymatic polymerization of 2-MeImpG.

Finally, we measured the pH dependence of the reaction and found that it fit well to a model with one group with a pK_a of 6.9±0.1 requiring deprotonation and another group with a pK_a of 8.2±0.1 requiring protonation (Figure 2.6A). In comparison, Wu and Orgel have reported a constant rate for this reaction from pH 6.5 to 8.0, although they did not present their data (Wu and Orgel 1992a). The most likely position to require protonation is the 2-methylimidazole group of 2-MeImpG since protonation would turn it from a very

poor leaving group (pK_a >14) to a good leaving group (pK_a = 7.9). The pK_a of the 2-methylimidazole nitrogen in 2-MeImpG has previously been estimated at 7.7 (Kanavarioti et al. 1989), but it could be raised near the polyanionic primer-template complex. For comparison, we measured the pH dependence of the reaction with ImpG, which has a pK_a of about 6.0 (Kanavarioti et al. 1989), and fit pK_a values of 6.6±0.1 and 9.6±0.1 with rates less than one tenth those of 2-MeImpG (Figure 2.6D). It is likely that ImpG follows a different mechanism and the pK_a of 9.6±0.1 represents the guanine nucleobase, although this mechanism still does not involve general acid-base catalysis in the rate-determining step (Figure 2.5B). Interestingly, the reactivity of pyrophosphate (Rohatgi, Bartel, and Szostak 1996) and 1-hydroxy-7-azabenzotriazole (Vogel, Deck, and Richert 2005) leaving groups, which do not require protonation, simply increases steadily with pH.

The lower pK_a is more difficult to explain. Although the 3′-hydroxyl group of the primer is much more nucleophilic when deprotonated, its pK_a is around 12.4 (Izatt et al. 1965). Mg^{2+} coordination can lower the pK_a of hydroxyl groups, for example, it lowers the pK_a of water from 15.7 to 11.4 (Dahm, Derrick, and Uhlenbeck 1993), and DNA polymerases use this effect to activate the 3′-hydroxyl (Steitz 1998). To test for Mg^{2+} activation in our system we measured the pH dependence with a 2′-deoxy terminated primer (Figure 2.6B), which should have raised the pK_a by up to 3 units (Izatt et al. 1965), and with Ca^{2+} instead of Mg^{2+} (Figure 2.6C), which should have raised the pK_a by 1.4 units (Dahm, Derrick, and Uhlenbeck 1993), but saw only minor changes to the observed pK_a values (Table 1). An alternative explanation is that at low pH the cytosine residues of one template are protonated and they form a triple helix with an unprotonated template bound to 2-MeImpG monomers. This phenomenon has been used to explain a drop in the rate of 2-

MeImpG polymerization on high molecular weight poly(C) templates below pH 7 (Inoue and Orgel 1982), but it typically requires higher concentrations of longer oligonucleotides and a lower pH than those used in this work (Manzini et al. 1990).

The pH dependence could also be interpreted as the same group requiring protonation and deprotonation at different steps in the reaction or in different positions, resulting in a peak rate near the pK_a of that group (Breslow 1991). In particular, Wu and Orgel found that the rate of addition of the first nucleotide in primer extension depends on the leaving group at the second position (Wu and Orgel 1992b), suggesting an interaction between the two that could require only one to be protonated. This possibility will be discussed extensively in Chapter 3.

In any case, a model has emerged where the majority of the rate enhancement from Mg^{2+} comes from coordination to the active site, but it does not activate the 3′-hydroxyl group through general base catalysis or by allowing it to be deprotonated prior to attack. Although hydroxyl groups are not very good nucleophiles, protonated 2-methylimidazole is a good leaving group. Hydrolysis of imidazole-activated phosphates has a highly negative β_{lg} value (Ruzicka and Frey 1993) indicating a loose transition state with significant bond breaking to the imidazole before nucleophilic attack. Indeed, given that the second order rate constant for hydrolysis of 2-MeImpG at pH 7.5 in the absence of Mg^{2+} is about 2.4×10^{-4} h^{-1} M^{-1} (Kanavarioti et al. 1989), the effective molarity of the 3′-hydroxyl to the 2-MeImpG phosphate would have to be about 10^4 to give the rates observed in this study. This value is within the range observed for other catalytic complexes (Cacciapaglia, Di Stefano, and Mandolini 2004). Therefore, orientation of the 3′-hydroxyl and 2-MeImpG phosphate and

stabilization of any negative charge build up on the phosphate in the transition state are sufficient to explain the observed rates (Figure 2.7).

This model could be tested further by using a phosphorothioate analog of 2-MeImpG, which would disrupt binding to Mg^{2+} and reveal if a Mg^{2+}-phosphate interaction is important for the reaction (Rajagopal, Doudna, and Szostak 1989). However, since the sulfur substitution would have to be on the reactive phosphate it would likely affect the reaction in other ways and make the results difficult to interpret. The interdependence between Mg^{2+} and 2-MeImpG binding should also be measured and compared to $Co(NH_3)_6^{3+}$ to better understand the dynamics of the primer-template-Mg^{2+}-2-MeImpG complex. Since the complex is so weak, simple molecules designed to stabilize it could substantially decrease the amount of Mg^{2+} and 2-MeImpG required to reach saturation. It should even be possible to develop a small molecule catalyst that can substitute for Mg^{2+} by specific hydrogen bonding to the 3′-hydroxyl and 2-MeImpG phosphate, similar to those developed for RNA hydrolysis based on mechanistic insights (Lönnberg 2011; Cheng, Abhilash, and Breslow 2012). If the requirement for divalent cation catalysis could be limited it would be possible to encapsulate non-enzymatic RNA polymerization in fatty acid vesicles, which has been a longstanding goal in origins of life and protocell research (Szostak 2012).

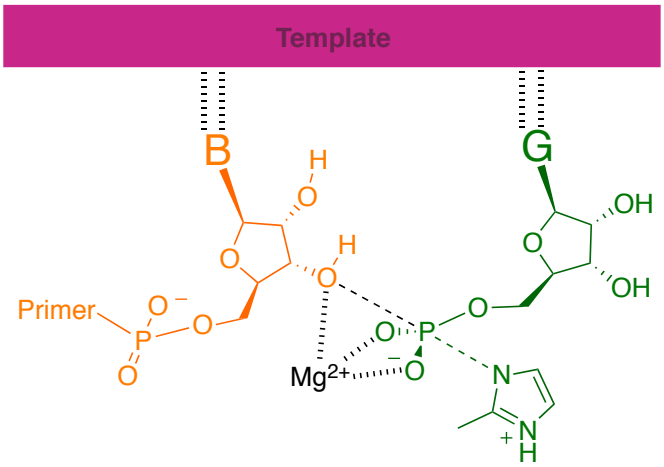


Figure 2.7. Model of the transition state for the primer extension reaction. Magnesium coordinates the primer 3′-hydroxyl and 2-MeImpG phosphate to orient them in a favorable geometry and stabilize negative charge build up on the phosphate.

Methods and Materials

General

All reagents were purchased from Sigma-Aldrich unless otherwise indicated. HPLC purified oligonucleotides were purchased from IDT DNA. All gels were imaged on a Typhoon Scanner 9400 (GE Healthcare) and analyzed using ImageQuant TL software (GE Healthcare). All further analysis was performed with Excel (Microsoft) and Prism (GraphPad).

Synthesis of guanosine 5′-phosphor-2-methylimidazolide

The activated nucleotides 2-MeImpG and ImpG were synthesized with slight modifications to the procedure of Joyce *et al.* (Joyce, Inoue, and Orgel 1984). Guanosine 5′-monophosphate free acid (MP Biomedicals) (210 mg) was mixed with 2-methylimidazole

(for 2-MeImpG) or imidazolide (for ImpG) (422 mg), dissolved in water and lyophilized to a dry solid. The solid was dissolved in anhydrous dimethylsulfoxide (40 mL) under argon and then triethylamine (0.21 mL), triphenylphosphine (294 mg) and 2,2′-dipyridyldisulfide (334 mg) were added. The reaction was stirred at room temperature for 23 h and then precipitated by dropwise addition to a flask containing acetone (400 mL), diethyl ether (250 mL), triethylamine (30 mL) and a solution of acetone saturated with sodium perchlorate (2 mL). The precipitate was collected by centrifugation, washed twice with 1:1 acetone-ether solution (40 mL) and purified by reverse phase column chromatography on a CombiFlash R_f system (Teledyne Isco) using a 50 g HP C18 Aq column and 20 mM triethylammonium bicarbonate pH 7.5 buffer with a linear gradient of 1% to 3% acetonitrile in the first 27 min and then up to 14% in the next 20 min at a flow rate of 15 mL/min. All eluate containing the product was combined and lyophilized at -10°C to a dry powder. The product was distributed into roughly 10 mg batches and was dissolved in water to a stock concentration of roughly 500 mM determined by UV spectroscopy assuming an extinction coefficient of 14.09 mM^{-1} cm^{-1} at 252 nm as per guanosine 5′-monophosphate (Cavaluzzi and Borer 2004). Stock solutions were stored at -80°C .

Primer extension reaction

Unless otherwise indicated, the conditions for the primer extension reactions were 200 mM HEPES pH 7.5, 50 mM $MgCl_2$, 50 mM 2-MeImpG, 0.3 μM Cy3-labeled RNA primer and 1.5 μM RNA C_4 template. The primer sequence was 5′-Cy3-GCG UAG ACU GAC UGG-3′ and the template sequence was 5′-AAC CCC CCA GUC AGU CUA CGC-3′. The total reaction volume was 30 μL and at each time point a 5 μL aliquot was taken, mixed with stop buffer

(100 mM EDTA, 6.4 M urea, 1X TBE buffer), and stored at -80°C. Samples were then ethanol precipitated and redissolved in loading buffer (7.2 M urea, 1X TBE buffer) before analysis by 20% urea-TBE PAGE. Gel intensities were fit to the following formula to get the pseudo-first order rate constant k_{obs}, where C is the y-intercept, and Primer/Primer$_0$ is the fraction of lane intensity that is primer at time t:

$$\ln(\text{Primer}/\text{Primer}_0) = -k_{obs} \times t + C \qquad (2.1)$$

Solvent deuterium kinetic isotope effect

A sample of HEPES powder was dissolved in D_2O and lyophilized and then redissolved in D_2O, titrated to a pH meter reading of 7.1 (indicating a pD of 7.5) with NaOD (Cambridge Isotopes Laboratories) and brought to 1 M. A similar buffer was made in H_2O and brought to a pH of 7.5 with NaOH. Reactions were prepared as indicated above without the HEPES buffer and then they were lyophilized and redissolved in the H_2O and D_2O buffers and finally 2-MeImpG was added to a final concentration of 60 mM from a stock solution in H_2O to begin each reaction. Using 2-MeImpG as a monomer, the volume percentages of D_2O used were 12.3, 25, 37.3, 50, 62.5, 75, and 87.7% corresponding to mole percentages of D of 0, 12.2, 24.8, 37.0, 49.5, 61.9, 74.3, and 86.9% according to published conversion factors (K. B. Schowen and Schowen 1982). Using ImpG as a monomer the volume percentages were 0, 23.2, 46.5, 69.7, and 92.3% corresponding to mole percentages of D of 0, 23, 46, 69, and 91.4%.

pH dependence

Reactions were performed according to the standard protocol, but with a buffer concentration of 500 mM to ensure the pH was not shifted by the 2-MeImpG. The following buffers were used: MES pH 5.5, 6.0 and 6.5, HEPES pH 6.5, 7.0, 7.5, 8.0 and 8.5, and CHES pH 8.54, 9.0, 9.5 and 10. The observed rate constants were fit to the following equation, where k_{max} is the rate constant when in the optimal protonation state, pK_{a1} is the pK_a of a group requiring deprotonation and pK_{a2} is the pK_a of a group requiring protonation:

$$k_{obs} = k_{max}\left(\frac{10^{pH-pK_{a1}}}{1+10^{pH-pK_{a1}}}\right)\left(1-\frac{10^{pH-pK_{a2}}}{1+10^{pH-pK_{a2}}}\right) \qquad (2.2)$$

References

Bevilacqua, Philip C. 2003. "Mechanistic Considerations for General Acid-Base Catalysis by RNA: Revisiting the Mechanism of the Hairpin Ribozyme." *Biochemistry* 42 (8) (March 4): 2259–2265. doi:10.1021/bi027273m.

Bowman, Jessica C, Timothy K Lenz, Nicholas V Hud, and Loren Dean Williams. 2012. "Cations in Charge: Magnesium Ions in RNA Folding and Catalysis." *Current Opinion in Structural Biology* 22 (3) (June 1): 262–272. doi:10.1016/j.sbi.2012.04.006.

Breslow, R. 1991. "How Do Imidazole Groups Catalyze the Cleavage of RNA in Enzyme Models and in Enzymes? Evidence From 'Negative Catalysis'." *Accounts of Chemical Research* 24 (11): 317–324.

Breslow, R. 1993. "Kinetics and Mechanism in RNA Cleavage." *Proceedings of the National Academy of Sciences USA* 90 (January 11): 1208–1211.

Cacciapaglia, Roberta, Stefano Di Stefano, and Luigi Mandolini. 2004. "Effective Molarities in Supramolecular Catalysis of Two-Substrate Reactions." *Accounts of Chemical Research* 37 (2) (February): 113–122. doi:10.1021/ar020076v.

Castro, Christian, Eric Smidansky, Kenneth R Maksimchuk, Jamie J Arnold, Victoria S Korneeva, Matthias Götte, William Konigsberg, and Craig E Cameron. 2007. "Two Proton Transfers in the Transition State for Nucleotidyl Transfer Catalyzed by RNA-

and DNA-Dependent RNA and DNA Polymerases." *Proceedings of the National Academy of Sciences USA* 104 (11) (March 13): 4267–4272. doi:10.1073/pnas.0608952104.

Cavaluzzi, Michael J, and Philip N Borer. 2004. "Revised UV Extinction Coefficients for Nucleoside-5'-Monophosphates and Unpaired DNA and RNA." *Nucleic Acids Research* 32 (1): e13. doi:10.1093/nar/gnh015.

Cheng, Liang, K G Abhilash, and Ronald Breslow. 2012. "Binding and Biomimetic Cleavage of the RNA Poly(U) by Synthetic Polyimidazoles." *Proceedings of the National Academy of Sciences USA* 109 (32) (August 7): 12884–12887. doi:10.1073/pnas.1210846109.

Dahm, S C, W B Derrick, and O C Uhlenbeck. 1993. "Evidence for the Role of Solvated Metal Hydroxide in the Hammerhead Cleavage Mechanism." *Biochemistry* 32 (48): 13040–13045.

Deck, Christopher, Mario Jauker, and Clemens Richert. 2011. "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA." *Nature Chemistry* 3 (8) (August): 603–608. doi:10.1038/nchem.1086.

Draper, David E, Dan Grilley, and Ana Maria Soto. 2005. "Ions and RNA Folding." *Annual Review of Biophysics and Biomolecular Structure* 34: 221–243. doi:10.1146/annurev.biophys.34.040204.144511.

Gilbert, W. 1986. "The RNA World." *Nature* 319 (6055): 618–618.

Horton, T E, and V J DeRose. 2000. "Cobalt Hexammine Inhibition of the Hammerhead Ribozyme." *Biochemistry* 39 (37) (September 19): 11408–11416.

Inoue, T, and L E Orgel. 1981. "Substituent Control of the PolyC-Directed Oligomerization of Guanosine 5'-Phosphorimidazolide." *Journal of the American Chemical Society* 103 (25): 7666–7667.

Inoue, T, and L E Orgel. 1982. "Oligomerization of (Guanosine 5'-Phosphor)-2-Methylimidazolide on Poly(C). an RNA Polymerase Model." *Journal of Molecular Biology* 162 (1) (November 25): 201–217.

Inoue, T, and L E Orgel. 1983. "A Nonenzymatic RNA Polymerase Model." *Science* 219 (4586) (February 18): 859–862.

Izatt, R M, L D Hansen, J H Rytting, and J J Christensen. 1965. "Proton Ionization From Adenosine." *Journal of the American Chemical Society* 87 (June 20): 2760–2761.

Johnston, WK, PJ Unrau, MS Lawrence, ME Glasner, and D P Bartel. 2001. "RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension." *Science* 292 (5520): 1319–1325.

Joyce, G F, T Inoue, and L E Orgel. 1984. "Non-Enzymatic Template-Directed Synthesis on

RNA Random Copolymers. Poly(C, U) Templates." *Journal of Molecular Biology* 176 (2) (June 25): 279–306.

Kanavarioti, A, C F Bernasconi, D J Alberas, and E E Baird. 1993. "Kinetic Dissection of Individual Steps in the Poly(C)-Directed Oligoguanylate Synthesis From Guanosine 5'-Monophosphate 2-Methylimidazolide." *Journal of the American Chemical Society* 115 (19): 8537–8546.

Kanavarioti, A, CF Bernasconi, D L Doodokyan, and D J Alberas. 1989. "Magnesium-Ion Catalyzed P-N Bond Hydrolysis in Imidazolide-Activated Nucleotides - Relevance to Template-Directed Synthesis of Polynucleotides." *Journal of the American Chemical Society* 111 (18): 7247–7257.

Kluger, R, and F H Westheimer. 1969. "The pH-Rate Profile for the Hydrolysis of Some Esters of a Bicyclic Phosphinic Acid. Evidence for Rate-Limiting Pseudorotation." *Journal of the American Chemical Society* 91 (August 30): 4143–4150.

Lassila, Jonathan K, Jesse G Zalatan, and Daniel Herschlag. 2011. "Biological Phosphoryl-Transfer Reactions: Understanding Mechanism and Catalysis." *Annual Review of Biochemistry* 80 (1) (July 7): 669–702. doi:10.1146/annurev-biochem-060409-092741.

Loeb, L A, and T A Kunkel. 1982. "Fidelity of DNA Synthesis." *Annual Review of Biochemistry* 51: 429–457. doi:10.1146/annurev.bi.51.070182.002241.

Lohrmann, R, P K Bridson, and L E Orgel. 1980. "Efficient Metal-Ion Catalyzed Template-Directed Oligonucleotide Synthesis." *Science* 208 (4451) (June 27): 1464–1465.

Lönnberg, Harri. 2011. "Cleavage of RNA Phosphodiester Bonds by Small Molecular Entities: a Mechanistic Insight." *Organic & Biomolecular Chemistry*: 1–17. doi:10.1039/c0ob00486c.

Lönnberg, Harri, R Strömberg, and A Williams. 2004. "Compelling Evidence for a Stepwise Mechanism of the Alkaline Cyclisation of Uridine 3'-Phosphate Esters." *Organic & Biomolecular Chemistry* 2 (15): 2165–2167. doi:10.1039/b406926a.

Manzini, G, L E Xodo, D Gasparotto, F Quadrifoglio, G A van der Marel, and J H van Boom. 1990. "Triple Helix Formation by Oligopurine-Oligopyrimidine DNA Fragments. Electrophoretic and Thermodynamic Behavior." *Journal of Molecular Biology* 213 (4) (June 20): 833–843. doi:10.1016/S0022-2836(05)80267-0.

Nesbitt, S, L A Hegg, and M J Fedor. 1997. "An Unusual pH-Independent and Metal-Ion-Independent Mechanism for Hairpin Ribozyme Catalysis." *Chemistry & Biology* 4 (8) (August): 619–630.

Perreault, D M, and E Anslyn. 1997. "Unifying the Current Data on the Mechanism of Cleavage - Transesterification of RNA." *Angewandte Chemie-International Edition in English* 36 (November 12): 432–450.

Prabahar, K J, and J P Ferris. 1997. “Adenine Derivatives as Phosphate-Activating Groups for the Regioselective Formation of 3′,5′-Linked Oligoadenylates on Montmorillonite: Possible Phosphate-Activating Groups for the Prebiotic Synthesis of RNA.” *Journal of the American Chemical Society* 119 (19) (May 14): 4330–4337.

Rajagopal, J, J A Doudna, and J W Szostak. 1989. “Stereochemical Course of Catalysis by the Tetrahymena Ribozyme.” *Science* 244 (4905) (May 12): 692–694.

Rohatgi, R, D P Bartel, and J W Szostak. 1996. “Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation.” *Journal of the American Chemical Society* 118 (14): 3332–3339.

Ruzicka, F J, and P A Frey. 1993. “The Ph-Dependence for the Hydrolysis of Uridine 5'-Phosphoimidazolates.” *Bioorganic Chemistry* 21 (2): 238–248.

Schowen, K B, and R L Schowen. 1982. “Solvent Isotope Effects of Enzyme Systems.” *Methods in Enzymology* 87: 551–606.

Shechner, David M, and D P Bartel. 2011. “The Structural Basis of RNA-Catalyzed RNA Polymerization.” *Nature Structural & Molecular Biology* 18 (9) (September): 1036–1042. doi:10.1038/nsmb.2107.

Steitz, TA. 1998. “Structural Biology - a Mechanism for All Polymerases.” *Nature* 391 (6664): 231–232.

Sulston, J, R Lohrmann, L E Orgel, and H T Miles. 1968. “Nonenzymatic Synthesis of Oligoadenylates on a Polyuridylic Acid Template.” *Proceedings of the National Academy of Sciences USA* 59 (3) (March): 726–733.

Szostak, J W. 2012. “The Eightfold Path to Non-Enzymatic RNA Replication.” *Journal of Systems Chemistry* 3 (September 7): 2.

Venkatasubban, K S, and R L Schowen. 1984. “The Proton Inventory Technique.” *CRC Critical Reviews in Biochemistry* 17 (1): 1–44.

Vogel, Stephanie R, Christopher Deck, and Clemens Richert. 2005. “Accelerating Chemical Replication Steps of RNA Involving Activated Ribonucleotides and Downstream-Binding Elements.” *Chemical Communications* (39) (October 21): 4922–4924. doi:10.1039/b510775j.

Wochner, Aniela, James Attwater, Alan Coulson, and Philipp Holliger. 2011. “Ribozyme-Catalyzed Transcription of an Active Ribozyme.” *Science* 332 (6026) (April 8): 209–212. doi:10.1126/science.1200752.

Wu, T F, and L E Orgel. 1992a. “Nonenzymatic Template-Directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides.” *Journal of the American Chemical Society* 114 (1): 317–322.

Wu, T F, and L E Orgel. 1992b. “Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 2. Templates Containing Cytidine and Guanosine Residues.” *Journal of the American Chemical Society* 114 (14): 5496–5501.

Wu, T F, and L E Orgel. 1992c. “Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 3. Incorporation of Adenosine and Uridine Residues.” *Journal of the American Chemical Society* 114 (21) (October 7): 7963–7969.

Yang, W, JY Lee, and M Nowotny. 2006. “Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity.” *Molecular Cell* 22 (1): 5–13. doi:10.1016/j.molcel.2006.03.013.

Chapter Three: Activating group interactions in non-enzymatic RNA polymerization

J. Craig Blain, Isabel M. Vogt and Jack W. Szostak

Abstract

In non-enzymatic RNA primer extension reactions, subtle changes to the nucleotide activating group can have a profound effect on the efficiency of the reaction. It has previously been shown that the primer extension reaction is faster when the nucleotide bound downstream of the adding nucleotide is activated with 2-methylimidazole as compared to imidazole. We quantified this effect and found that the rate was 2.7 times faster with a downstream 2-methylimidazole-activated nucleotide, whereas the phosphor-2-methylimidazolide reacted only 2.0 times faster than the phosphorimidazolide when both had a phosphor-2-methylimidazolide bound downstream. The difference could not be attributed to weaker binding by imidazole-activated nucleotides, since we found that they saturated the reaction rate at lower concentrations. In the course of studying this phenomenon, we found that free 2-methylimidazole inhibited primer extension with guanosine 5′-phosphor-2-methylimidazolide (2-MeImpG). Weaker inhibitory effects were observed for 2-methylimidazole on the hydrolysis of 2-MeImpG and imidazole on primer extension with guanosine 5′-phosphorimidazolide. The inhibition could not be explained by competitive binding to magnesium, since an interaction between 2-methylimidazole and magnesium was not observed by NMR. We favor a model where the activating groups of the reacting and downstream binding nucleotides form a favorable interaction, and free activating group disrupts that interaction. The implications for this model on non-enzymatic RNA replication and the origin of life are discussed.

Introduction

One of the earliest stages in the development of life likely involved the self-replication of RNA sequences through template-directed, non-enzymatic polymerization (Joyce and Orgel 2006). For this process to be sustainable, the template copying reaction must be efficient and rapid enough to compete with RNA hydrolysis and other side reactions; however, the copying reaction requires the condensation of nucleotides in water to form large molecules with relatively low entropy and it is therefore energetically unfavorable. As a result, the 5′-phosphate of the nucleotides must be activated either *in situ* or prior to the reaction. In the late 1960s Leslie Orgel and colleagues began a decades long research program that laid the foundation for understanding these reactions.

The pyrophosphate group used in modern biology reacts far too slowly to be used in non-enzymatic polymerization reactions (Rohatgi, Bartel, and Szostak 1996; Orgel 2004). The first successful non-enzymatic template copying reactions used condensing reagents to activate nucleotides *in situ*. In particular, the water soluble carbodiimide 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDAC) was used to activate adenosine 5′-monophosphate for polymerization on a poly(U) template (Sulston et al. 1968). However, these reactions were inefficient and condensing reagents can alkylate nucleobases (Gilham 1962; Chu, Wahl, and Orgel 1983). It was quickly discovered that longer polymers were obtained if the nucleotide was pre-activated with imidazole (Weimann et al. 1968). Interestingly, both EDAC and imidazole activation lead to the preferential formation of 2′-5′ linkages in the backbone using Mg^{2+} as a catalyst, whereas Zn^{2+} catalysis gave the natural 3′-5′ linkage with imidazole-activated monomers (Lohrmann, Bridson, and Orgel 1980). To further improve the reaction yields, Inoue and Orgel compared imidazole to its 2-methyl, 2-

ethyl and 4-methyl derivates and found that 2-methylimidazole-activated nucleotides polymerize substantially faster than the others and form 3′-5′ linkages even with Mg^{2+} as a catalyst (Inoue and Orgel 1981).

It is still not understood what determines how good an activating group will be and few other groups have been found that are as good as 2-methylimidazole (Kaiser and Richert 2013). There are presumably at least three important factors: steric effects, the pK_a of the activating group, and the protonation state of the activating group if applicable (see Chapter 2). The imidazole activating group is much more reactive when protonated since protonation changes the effective leaving group from negatively charged imidazolide (pK_a > 14) to neutral imidazole ($pK_a \sim 7$). 2-Methylimidazole has a higher pK_a at 7.9, which should make it a worse leaving group, but it also allows it to be protonated at a higher pH. Consistent with this analysis, Kanavarioti *et al.* found that guanosine 5′-phosphorimidazolide (ImpG) hydrolyses more quickly than 5′-phosphor-2-methylimidazolide (2-MeImpG) at low pH, but 2-MeImpG hydrolyses more quickly from pH 7.5 to 9 (Kanavarioti et al. 1989). We have also found that the ability of 2-MeImpG to be protonated at a higher pH could explain why it is more reactive in primer extension reactions (see Chapter 2).

When Wu and Orgel analyzed non-enzymatic polymerization more thoroughly using primer extension reactions they noticed another important factor: the rate of addition of a nucleotide is affected by the activating group of the nucleotide it is stacked against (Figure 3.1) (Wu and Orgel 1992). In particular, a 2-MeImpC monomer adds to a primer much more quickly when the next nucleotide downstream is 2-MeImpG than it does when that nucleotide is ImpG. This neighboring activating group effect has important implications for

the optimization of the activating group and for conditions and catalysts for the polymerization reaction. Therefore, we studied some of the potential causes of this effect and the implications it could have for *in situ* nucleotide reactivation.

Figure 3.1. Model of guanosine 5′-monophosphate monomers (green) activated with 2-methylimidazole (red) bound to an RNA template (magenta) next to a primer (orange). The activating group of the neighboring monomer influences the phosphoryl transfer reaction between the primer and the first monomer. Based on an idealized RNA duplex structure.

Results

Quantitation of the neighboring activating group effect on primer extension

We first sought to confirm and quantify the effect of the neighboring activating group reported by Wu and Orgel (Wu and Orgel 1992). To do so, we used the primer extension assay discussed in Chapter 2 (Figure 2.2), but with a 3′-GCCC-5′ template, 2-

MeImpC as the first monomer added, and 2-MeImpG or ImpG as the stacking monomer. To compensate for the total monomer concentration of 100 mM (50 mM each) we also increased the $MgCl_2$ concentration to 100 mM. Consistent with the observations of Wu and Orgel, the rate constant with ImpG as the stacking monomer was only 37% that with 2-MeImpG (Figure 3.2A). For comparison, we performed the same experiment, but with a 3′-CGGG-5′ template so that the first monomer added was either 2-MeImpG or ImpG, and it was always stacked to 2-MeImpC (Figure 3.2B). We found that the rate constant for ImpG extension was 50% that of 2-MeImpG extension, indicating that the neighboring activating group is at least as important for the rate of addition as the activated group of the reacting monomer. In each case, very little primer extension was observed when the stacking monomer was omitted entirely.

Since we only used a single concentration for each monomer in these experiments, it was important to rule out the possibility that ImpG simply bound less tightly to the template than 2-MeImpG. We tested this possibility by measuring the dependence of the reaction rate on the ImpG concentration to determine the saturation point (Figure 3.2C). We found that the plot fit well to a single binding site model with a K_D of 5±2 mM (estimate ± standard error) and a k_{max} of 0.15±0.01 h^{-1}, thus ImpG saturated at a lower concentration than 2-MeImpG (Figure 2.3). Therefore, low occupancy does not explain the lower rate with ImpG.

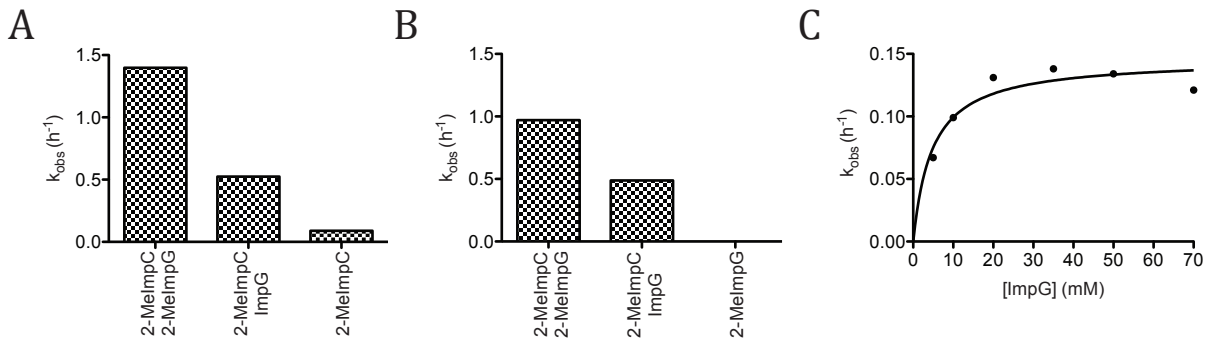


Figure 3.2. Effect of the neighboring monomer's activating group on the rate of primer extension. The rate constant for the first addition was measured using the indicated monomers and a (A) 3′-GCCC-5′ or (B) 3′-CGGG-5′ template. (C) The dependence of the rate constant on the concentration of ImpG monomer with a C_4 template.

Inhibition of non-enzymatic polymerization by 2-methylimidazole and imidazole

One possible role for the neighboring activating group is to make an important interaction that stabilizes the complex in a reactive geometry or to help catalyze the reaction. If that were the case, free 2-methylimidazole could inhibit or enhance the reaction, respectively, by competing for the same interaction. We measured the rate constant for primer extension with 2-MeImpG on a C_4 template in the presence of varying concentrations of free 2-methylimidazole and found that it inhibited the reaction, with 10 mM 2-methylimidazole decreasing the rate by more than one half (Figure 3.3A). Similarly, primer extension reactions with an ImpG monomer were inhibited by free imidazole, although the inhibition was less efficacious (Figure 3.3B).

We next sought to determine if the inhibitory effect of free 2-methylimidazole was competitive, uncompetitive or noncompetitive by measuring the dependence of the rate on the concentration of 2-MeImpG at different concentrations of 2-methylimidazole (Figure

3.3C). Monomer self-inhibition made it difficult to rigorously fit the data; however, increasing concentrations of 2-methylimidazole seemed to decrease the maximum achievable rate while the effect on the apparent K_D was unclear. This trend is consistent with 2-methylimidazole disrupting an interaction made by the activating groups in the reactive complex.

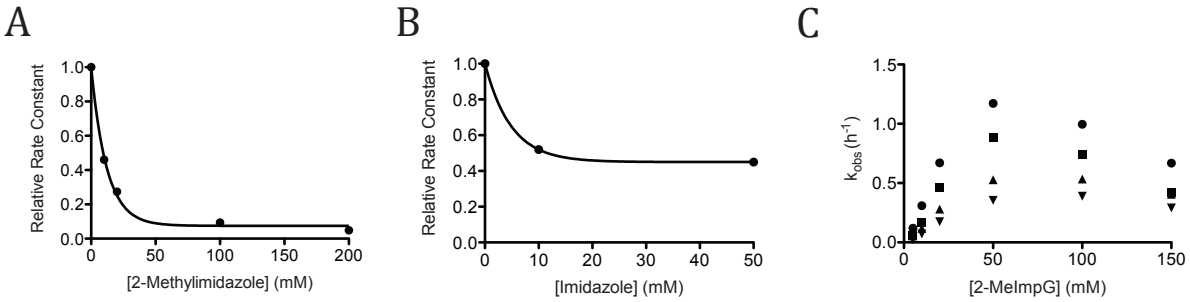


Figure 3.3. Inhibition of primer extension by free activating group. (A) Effect of free 2-methylimidazole on the rate of primer extension of 2-MeImpG. (B) Effect of free imidazole on the rate of primer extension of ImpG. (C) Dependence of the rate of primer extension on the concentration of 2-MeImpG with 0 mM (•), 5 mM (■), 10 mM (▲), and 20 mM (▼) 2-methylimidazole. All reactions used a C_4 template.

NMR analysis of a possible 2-methylimidazole-Mg^{2+} interaction

The mechanism of catalysis by Mg^{2+} could involve the coordination of Mg^{2+} to the 2-methylimidazole nitrogen lone pair of a 2-MeImpG monomer. If that were the case, then free 2-methylimidazole could compete with 2-MeImpG for that interaction and disrupt catalysis. We tested for a direct interaction between Mg^{2+} and 2-methylimidazole by measuring the effect of $MgCl_2$ on the ^1H NMR chemical shift of the aromatic protons of 10

mM 2-methylimidazole (Figure 3.4). Using the formyl proton of *N*,*N*-dimethylformamide (DMF) as a reference, we found that $MgCl_2$ had a minimal effect up to a concentration of 100 mM. Therefore, 2-methylimidazole and Mg^{2+} do not interact at concentrations relevant to the primer extension reaction.

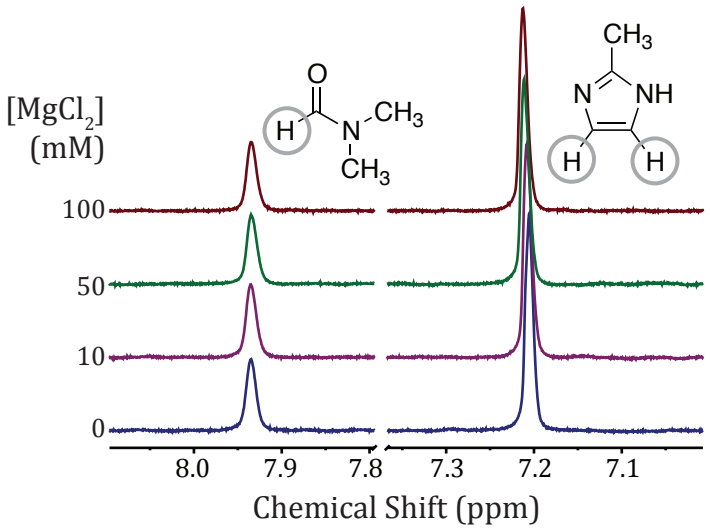


Figure 3.4. Lack of interaction between 2-methylimidazole and Mg^{2+}. The ^1H NMR spectrum of 10 mM 2-methylimidazole, with 10 mM DMF as a reference, was measured in the presence of increasing concentrations of $MgCl_2$. The $MgCl_2$ had did not affect the chemical shift of 2-methylimidazole relative to DMF (change < 0.01 ppm) indicating a lack of interaction.

Effect of 2-methylimidazole on the hydrolysis of 2-MeImpG and 2-MeImpC

The hydrolysis of activated nucleotides is chemically very similar to primer extension, but it does not involve the formation of a reactive complex. Therefore, the effect of 2-methylimidazole on the rate of hydrolysis can be used as a proxy for its effect on the

phosphoryl transfer chemistry alone. We measured the hydrolysis rate of 5 mM 2-MeImpG at pH 7.5 and 20°C by ^{31}P NMR in the presence and absence of 200 mM $MgCl_2$ and 100 mM 2-methylimidazole (Table 3.1). The rate of 2-MeImpG hydrolysis in the absence of $MgCl_2$ was slow, with a pseudo-first order rate constant of 5.5±0.2 $x10^{-3}$ h^{-1}, corresponding to a half-life of 5.3±0.2 days, and the presence of $MgCl_2$ increased the rate by a factor of 1.6. The addition of 2-methylimidazole inhibited hydrolysis, decreasing the uncatalyzed and $MgCl_2$-catalyzed rates by a factor of 1.7 and 2.0, respectively. To test if an interaction between the nucleobase and 2-methylimidazole was important for inhibition, we measured the effect of 2-methylimidazole on the rate of uncatalyzed 2-MeImpC hydrolysis and found that it similarly decreased the rate by a factor of 2.0. Therefore, 2-methylimidazole does inhibit the hydrolysis of phosphor-2-methylimidazolides, but less potently than it inhibits the primer extension reaction, for which a concentration of 10 mM was sufficient to halve the rate.

Table 3.1. Effect of $MgCl_2$ and 2-methylimidazole on the rate of hydrolysis of 5 mM activated nucleotide at pH 7.5 and 20°C measured by ^{31}P NMR. Values are presented as estimate ± standard error.

Monomer	[$MgCl_2$] (mM)	[2-Methylimidazole] (mM)	k_{obs} x 10^3 (h^{-1})	Half-life (d)
2-MeImpG	0	0	5.5±0.2	5.3±0.2
2-MeImpG	200	0	8.9±0.2	3.24±0.07
2-MeImpG	0	100	3.25±0.06	8.9±0.2
2-MeImpG	200	100	4.5±0.3	6.5±0.4
2-MeImpC	0	0	5.3±0.1	5.5±0.2
2-MeImpC	0	100	2.64±0.05	10.9±0.2

Discussion

Consistent with a report from Wu and Orgel (Wu and Orgel 1992), we found that the addition of an activated nucleotide to a primer is affected by the activating group of the neighboring nucleotide stacked against the adding nucleotide. In particular, a 2-MeImpC monomer will add to a primer 2.7 times more quickly when stacked to 2-MeImpG than when stacked to ImpG (Figure 3.2A). The effect of the stacking nucleotide's activating group was actually greater than that of the adding nucleotide, with 2-MeImpG adding 2.0 times more quickly than ImpG when stacked to 2-MeImpC (Figure 3.2B). The difference could not be explained by weak binding of the ImpG monomer (Figure 3.2C), indicating that the stacking nucleotide's activating group must be making an important interaction.

Furthermore, we found that free 2-methylimidazole inhibited the primer extension reaction with 2-MeImpG (Figure 3.3A). To the best of our knowledge, inhibition by free

activating group has not been observed in non-enzymatic RNA polymerization and it could have important implications for the origin of life. A number of possible mechanisms of inhibition by 2-methylimidazole are possible, including disrupting the interaction made with the stacking nucleotide's activating group, competing with 2-MeImpG for binding to the primer-template complex, competing with 2-MeImpG for binding to Mg^{2+}, and directly inhibiting the phosphoryl transfer reaction. We found that 2-methylimidazole inhibition could not be overcome by increasing the concentration of 2-MeImpG, indicating that it is not a competitive inhibitor of monomer binding (Figure 3.3C). However, the self-inhibitory effect of the monomer could have masked a rate recovery. We then tested for 2-methylimidazole-Mg^{2+} binding by ^{1}H NMR, but did not observe an effect of Mg^{2+} on the spectrum of 2-methylimidazole (Figure 3.4). Finally, we tested for a direct effect on the phosphoryl transfer chemistry by measuring the effect of 2-methylimidazole on the hydrolysis rate of 2-MeImpG (Table 3.1). We found that 100 mM 2-methylimidazole roughly halved the rate, whereas 10 mM was sufficient to halve the rate of primer extension.

From these results the mechanism of inhibition is difficult to determine. The imidazole group of activated nucleotides can exchange quickly with other imidazole derivatives (Kanavarioti et al. 1989), but it is unlikely that the exchange reaction goes through an intermediate that is stable enough to sequester the monomer from hydrolysis. One possible model is that the 2-methylimidazole group of the stacking 2-MeImpG monomer interacts directly with that of the adding monomer and that interaction directs the phosphate of the adding monomer into a favorable geometry for the reaction. Free 2-methylimidazole competes for this interaction and directs the 2-MeImpG to adopt an

unreactive geometry. The two 2-methylimidazole groups could associate through π-π, π-cation, hydrogen-bonding, or n-π* interactions. This model could be tested further by examining a series of activating groups in the stacking position to determine what properties of the group (*e.g.*, aromatic surface area, charge) are important. Similarly, these phenomena could be explored for the 1-hydroxy-7-azabenzotriazole (HOAt) activating group, which has also been used in non-enzymatic primer extension reactions (Deck, Jauker, and Richert 2011). We are currently using molecular dynamics simulations to gain a better understanding of the possible interactions and how they would affect the geometry of the system. It will also be interesting to see if it is the systems with an activating group interaction that show a preference for forming 3′-5′ linkages, suggesting that it directs them into an orientation for attack specifically by the 3′-hydroxyl.

In the origin of life, nucleotide activation and non-enzymatic polymerization likely occurred in the same environment. Therefore, the activating group would have to be present at fairly high concentrations during the primer extension reaction. Phosphor-2-methylimidazolides and any other nucleotides that display activating group inhibition would not perform well in these conditions and may be poor models for early non-enzymatic replication as a result. Even in the laboratory, methods for *in situ* nucleotide activation will be important in the development of continuously replicating RNA systems (Szostak 2012; Deck, Jauker, and Richert 2011). The very property that has made 2-methylimidazole so effective to this point might limit its future applications. Alternatively, the activating group interaction could be exploited to enhance the reaction. Bidentate activating groups could be used to fix a 2-MeImpG dimer in the reactive conformation, or catalysts could be developed to stabilize the interaction.

Methods and Materials

Synthesis of 2-MeImpG, ImpG and 2-MeImpC

All three activated nucleotides were synthesized as previously described (Chapter 2).

Primer extension reactions

Primer extension reactions were performed as previously described (Chapter 2), with a few minor changes. The primer and template concentrations were 0.6 and 3 μM, respectively. At each time point a 2.5 μL aliquot was taken, mixed with 7.5 μL loading buffer (7.2 M urea, 1X TBE buffer) and run directly on the denaturing PAGE gel. The GC_3 template had the sequence 5′-AAC CCG CCA GUC AGU CUA CGC-3′, and the CG_3 template had the sequence 5′-AAG GGC CCA GUC AGU CUA CGC-3′.

Hydrolysis of 2-MeImpG and 2-MeImpC by ^{31}P NMR

The hydrolysis of the activated nucleotides was measured under the following conditions: 5 mM activated nucleotide, 100 mM HEPES pH 7.5, 5 mM potassium phosphate pH 7.5, 10% (v/v) D_2O, and optionally 100 mM 2-methylimidazole pH 7.5 and 200 mM $MgCl_2$ at 20°C. The final reaction volume was 500 μL. The reactions were followed by ^{31}P NMR on a 400 MHz Varian spectrometer (Oxford AS-400) using 128 scans and spectra were analyzed using iNMR software (Nucleomatica), with time points taken daily for at least 12 days. At each time point the integration of the phosphorus peak for the activated nucleotide (2-MeImpG: -10.40 ppm; 2-MeImpC: -10.32 ppm) relative to the phosphate peak

(0 ppm) was measured and normalized to the initial integration. These values were fit to a first order exponential decay to get a pseudo-first order rate constant for the reaction.

Interaction between Mg^{2+} and 2-methylimidazole

The potential interaction between Mg^{2+} and 2-methylimidazole was studied under the following conditions: 10 mM 2-methylimidazole pH 7.5, 10 mM DMF, 200 mM HEPES pH 7.5, 55% (v/v) D_2O, and 0, 10, 50, and 100 mM $MgCl_2$ to a final volume of 500 μL with water at 25°C. The ^1H NMR spectrum of each sample was taken on a 400 MHz Varian spectrometer (Oxford AS-400) using 32 scans and spectra were analyzed using iNMR software (Nucleomatica). For the 0 mM $MgCl_2$ sample, the water peak was set to 4.79 ppm and the DMF formyl peak was measured to be 7.935 ppm. For all other samples, the DMF formyl peak was set to 7.935 ppm.

References

Chu, B C, G M Wahl, and L E Orgel. 1983. "Derivatization of Unprotected Polynucleotides." *Nucleic Acids Research* 11 (18) (September 24): 6513–6529.

Deck, Christopher, Mario Jauker, and Clemens Richert. 2011. "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA." *Nature Chemistry* 3 (8) (August): 603–608. doi:10.1038/nchem.1086.

Gilham, P T. 1962. "An Addition Reaction Specific for Uridine and Guanosine Nucleotides and Its Application to the Modification of Ribonuclease Action." *Journal of the American Chemical Society* 84 (October 17): 687–688.

Inoue, T, and L E Orgel. 1981. "Substituent Control of the PolyC-Directed Oligomerization of Guanosine 5'-Phosphorimidazolide." *Journal of the American Chemical Society* 103 (25): 7666–7667.

Joyce, G F, and L E Orgel. 2006. "Chapter 2: Progress Toward Understanding the Origin of the RNA World." In. CSHL Press.

Kaiser, Andreas, and Clemens Richert. 2013. “Nucleotide-Based Copying of Nucleic Acid Sequences Without Enzymes.” *The Journal of Organic Chemistry* 78 (3) (February 1): 793–799. doi:10.1021/jo3025779.

Kanavarioti, A, CF Bernasconi, D L Doodokyan, and D J Alberas. 1989. “Magnesium-Ion Catalyzed P-N Bond Hydrolysis in Imidazolide-Activated Nucleotides - Relevance to Template-Directed Synthesis of Polynucleotides.” *Journal of the American Chemical Society* 111 (18): 7247–7257.

Lohrmann, R, P K Bridson, and L E Orgel. 1980. “Efficient Metal-Ion Catalyzed Template-Directed Oligonucleotide Synthesis.” *Science* 208 (4451) (June 27): 1464–1465.

Orgel, L E. 2004. “Prebiotic Chemistry and the Origin of the RNA World.” *Critical Reviews in Biochemistry and Molecular Biology* 39 (2): 99–123. doi:10.1080/10409230490460765.

Rohatgi, R, D P Bartel, and J W Szostak. 1996. “Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation.” *Journal of the American Chemical Society* 118 (14): 3332–3339.

Sulston, J, R Lohrmann, L E Orgel, and H T Miles. 1968. “Nonenzymatic Synthesis of Oligoadenylates on a Polyuridylic Acid Template.” *Proceedings of the National Academy of Sciences USA* 59 (3) (March): 726–733.

Szostak, J W. 2012. “The Eightfold Path to Non-Enzymatic RNA Replication.” *Journal of Systems Chemistry* 3 (September 7): 2.

Weimann, B J, R Lohrmann, L E Orgel, H Schneider-Bernloehr, and J E Sulston. 1968. “Template-Directed Synthesis with Adenosine-5'-Phosphorimidazolide.” *Science* 161 (3839) (July 26): 387.

Wu, T F, and L E Orgel. 1992. “Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 2. Templates Containing Cytidine and Guanosine Residues.” *Journal of the American Chemical Society* 114 (14): 5496–5501.

Chapter Four: Synthesis and non-enzymatic polymerization of 2′-NH_2-TNA

J. Craig Blain, Alonso Ricardo and Jack W. Szostak

Abstract

Threose nucleic acid (TNA) is a promising alternative genetic material that may have played a role in the early evolution of life. In this work we developed a novel synthesis of TNA nucleosides with a 2′-amino modification (2′-NH_2-TNA). The synthesis was based on a cycloaddition reaction between a glycal and azodicarboxylate, followed by direct nucleosidation of the cycloadduct. Using this route we synthesized the thymine and guanine 2′-NH_2-TNA nucleosides in 7 steps with 24% and 12% overall yield, respectively. We then activated the guanine nucleoside with a 3′-phosphor-2-methylimidazolide group and tested its ability to copy C_4 RNA, DNA and TNA templates by non-enzymatic primer extension. We measured pseudo-first order rate constants for the first nucleotide addition step of 1.5±0.1, 0.97±0.06 and 0.57±0.06 h^{-1} on the RNA, DNA and TNA template, respectively, at pH 7.5 and 20°C with 5 mM activated nucleotide. These rates are more than an order of magnitude slower than those for other amino-modified ribose systems under the same conditions, and copying of the TNA template, which best represented a true self-copying reaction, was the slowest of all. Although its rate of addition is slow, the activated 2′-NH_2-TNA nucleotide is relatively stable, with a half-life of hydrolysis of 6.47±0.07 days. Given its other desirable properties, 2′-NH_2-TNA has potential as a self-replicating genetic material.

Introduction

The early development of life must have proceeded from abiotic chemicals, through simple cells capable of Darwinian evolution to the complex life forms that exist today. To better understand this process efforts are underway to synthesize model protocells composed of a self-replicating compartment and genetic material (Szostak, Bartel, and Luisi 2001). Simple systems of growth and division have been developed for fatty acid vesicles (Zhu and Szostak 2009). However, despite recent advances in autocatalytic RNA synthesis (Vaidya et al. 2012; Lincoln and Joyce 2009) and ribozyme catalyzed RNA dependent RNA polymerization (Wochner et al. 2011), the development of a self-replicating nucleic acid system capable of open-ended evolution has not been developed.

As an alternative to ribozyme-mediated replication, activated nucleotides can be used to copy sequences without enzymatic catalysis. Building on the pioneering work of Leslie Orgel and colleagues (Inoue and Orgel 1983; Lohrmann and Orgel 1976; T. F. Wu and Orgel 1992), we and others have been developing systems of non-enzymatic, template-directed primer extension using amino-modified nucleotides as models for genetic replication since they show enhanced rates of polymerization due to the greater nucleophilicity of the amine (Kaiser and Richert 2013; Kaiser et al. 2012; Röthlingshöfer et al. 2008; Chen, Cai, and Szostak 2009; Schrum et al. 2009; Zhang et al. 2013; Mansy et al. 2008). In particular, we have studied the polymerization of the acyclic 2′-amino-2′-deoxyglycerol nucleotides (Chen, Cai, and Szostak 2009), 2′-amino-2′,3′-dideoxyribose nucleotides (Schrum et al. 2009; Mansy et al. 2008), and 3′-amino-2′,3′-dideoxyribose nucleotides (Zhang et al. 2013). These systems are attractive as models of genetic self-replication since they have the potential to efficiently copy all four nucleobases on the

order of minutes (Zhang et al. 2013) and they are compatible with fatty acid vesicles (Mansy et al. 2008).

(L)-α-Threose nucleic acid (TNA) (Figure 4.1A), based on the four carbon sugar threose, has also generated a lot of interest since it was discovered by Albert Eschenmoser and colleagues that it is able to form stable Watson-Crick duplexes with itself, DNA and RNA (Eschenmoser 1999; Schoning et al. 2000; Schoning et al. 2002). Since TNA could have been synthesized by the same potentially prebiotic processes that have been proposed for RNA (Kim et al. 2011; Islam et al. 2013), it may have played a role in the early evolution of life (Orgel 2000). To support this hypothesis, TNA has been used as a template for the non-enzymatic polymerization of activated ribonucleotides (Heuberger and Switzer 2006) and *in vitro* selection techniques have been used to isolate functional TNA sequences (J. Ichida, Zou, et al. 2005; Yu, Zhang, and Chaput 2012) based on the enzymatic copying of DNA to TNA and *vice versa* (Chaput and Szostak 2003; Chaput, Ichida, and Szostak 2003; J. Ichida, Horhota, et al. 2005; Pinheiro et al. 2012; Yu et al. 2013).

Structurally, since TNA has a shortened repeating backbone unit the interphosphate distance in a TNA-TNA duplex is only 5.9 Å (Ebert et al. 2008). As a result, it has a shallow and wide minor groove, similar to A-form DNA and RNA (Ebert et al. 2008), and binds more tightly to RNA than DNA (Pallan et al. 2003). TNA residues adopt a unique 4′-*exo* sugar pucker with the 2′- and 3′-hydroxyl groups axial, which maximizes the interphosphate distance (Pallan et al. 2003; Wilds et al. 2002). Since it has only four rotatable bonds per residue, as opposed to five in RNA, the entropic cost of trapping a single-stranded TNA oligonucleotide into a duplex is less than with RNA or DNA (X. Wu, Guntha, et al. 2002; Ferencic et al. 2004). As a result, we hypothesized that single stranded TNA might form a

structure that is pre-orgranized for favorable nucleotide binding. Template pre-organization has previously been shown to enhance the rate of non-enzymatic polymerization for altritol (ANA) and hexitol (HNA) nucleic acids (Kozlov et al. 1999; Kozlov et al. 2000).

Given the desirable properties of TNA and its potential importance in the origin of life, we became interested in testing the activated nucleotides of 2′-amino-2′-deoxythreose nucleic acid (2′-NH_2-TNA) in non-enzymatic primer extension experiments. The Eschenmoser group has synthesized the adenine and thymine residues of the 2′- and 3′-amino-modified versions of TNA and shown that they have similar base-pairing properties to regular TNA, with the 2′-amino derivative forming significantly weaker duplexes (X. Wu, Guntha, et al. 2002; Ferencic et al. 2004). They have also shown that the amino modification enhances the rate of non-enzymatic ligation of two TNA oligonucleotides on a TNA template (X. Wu, Delgado, et al. 2002). Conveniently, the rigid threose ring would also hold the activated phosphate and nucleophilic amine of a monomer apart from each other to minimize cyclization, which can be rapid with amino nucleotides (Chen, Cai, and Szostak 2009; Zhang et al. 2013) (Figure 4.1B).

In the route developed by Eschenmoser *et al.* the thymine 2′-NH_2-TNA nucleotide (**4.8t**) was synthesized indirectly through the thymine TNA nucleoside and the adenine 2′-NH_2-TNA nucleoside was synthesized through trans-nucleosidation of **4.8t** (Ferencic et al. 2004). We first sought to develop a direct route for the synthesis of the guanine 2′-NH_2-TNA nucleoside (**4.8g**). We based our synthesis on a [4+2] cycloaddition reaction between an azodicarboxylate and glycal that was previously developed to synthesize *trans* 2-amino, 3-hydroxyl carbohydates (Fitzsimmons, Leblanc, and Rokach 1987; Leblanc and

Fitzsimmons 1989; Leblanc et al. 1989) due to its high efficiency and stereo- and regioselectivity. We then synthesized the 3′-phosphor-2-methylimidazolide derivative (2-MeImpntG, **4.11**) and tested it in non-enzymatic primer extension experiments on C_4 RNA, DNA and TNA templates. This work provides a novel synthetic approach to modified nucleosides and a better understanding of the potential role threose-based nucleic acids could play as a genetic material for a model protocell.

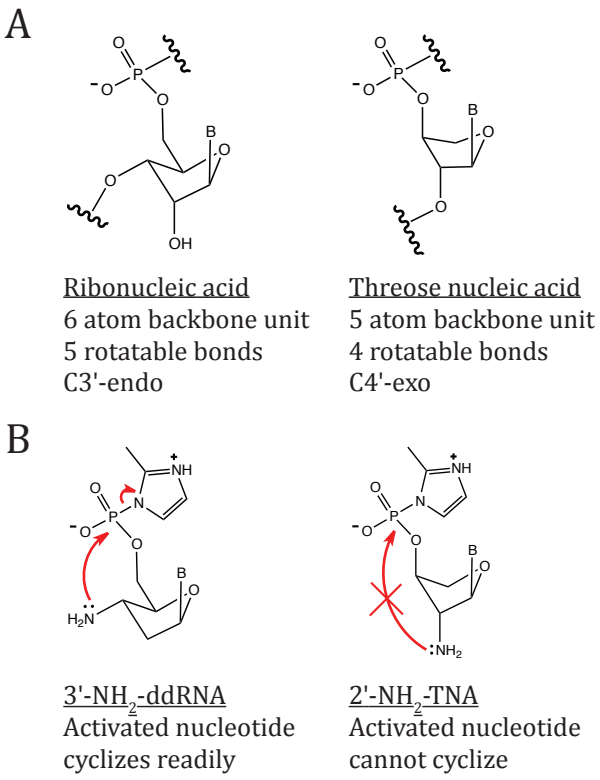


Figure 4.1. Structure of threose nucleic acids. (A) Comparison of the structure of RNA and TNA. (B) Comparison of the structure of activated 3′-amino-2′,3′-dideoxyribose (3′-NH_2-ddRNA) nuceotides and 2′-amino-2′-deoxythreose (2′-NH_2-TNA) nucleotides.

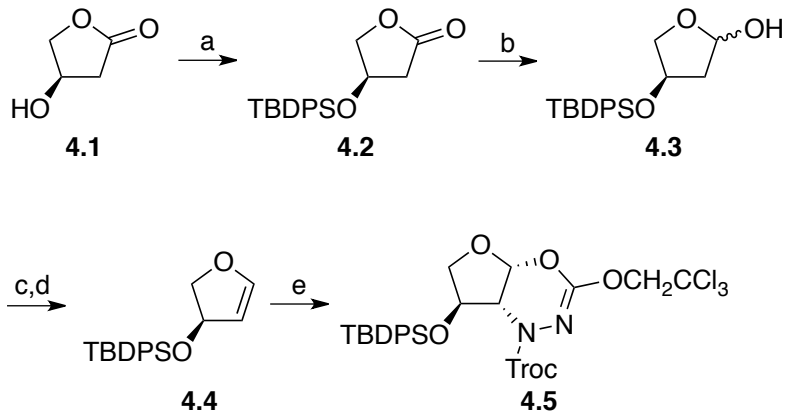
Results

Synthesis of 2′-NH_2-TNA nucleosides

We sought to develop an efficient synthesis that could yield multiple 2′-NH_2-TNA nucleosides from a common intermediate with high enantiomeric purity. To this end, we decided to base our synthesis on a [4+2] cycloaddition reaction between a glycal and azodicarboxylate that had previously been developed to synthesize 2-amino carbohydrates (Fitzsimmons, Leblanc, and Rokach 1987; Leblanc and Fitzsimmons 1989; Leblanc et al. 1989). Although this reaction had not been used to synthesize nucleosides or four carbon sugars, the general scheme was compatible with our needs.

We synthesized the glycal from (*R*)-β-hydroxy-γ-butyrolactone (**4.1**), which can be obtained in >99% e.e. As shown in Scheme 4.1, we first protected the hydroxyl group with a *tert*-butyldiphenylsilyl (TBDPS) group since it is bulky and non-electron withdrawing. Next, we reduced the lactone to lactol **4.3** with diisobutylaluminum hydride (DIBAL-H), converted the 1-hydroxy group to a good leaving group by mesylation and then eliminated it with heating to afford glycal **4.4**. These reactions were performed at a >5 g scale in high yield. For the cycloaddition reaction we mixed the glycal with bis(2,2,2-trichloroethyl) azodicaroboxylate (BTCEAD) in cyclohexane with constant 350 nm irradiation to allow conversion of the azodicarboxylate to the more reactive *cis* isomer (Gustorf et al. 1970). Previous reports have used dichloromethane as a co-solvent (Leblanc and Fitzsimmons 1989); however, we obtained higher yields in neat cyclohexane.

Scheme 4.1. Synthesis of the nucleosidation precursor by [4+2] cycloaddition.[a]

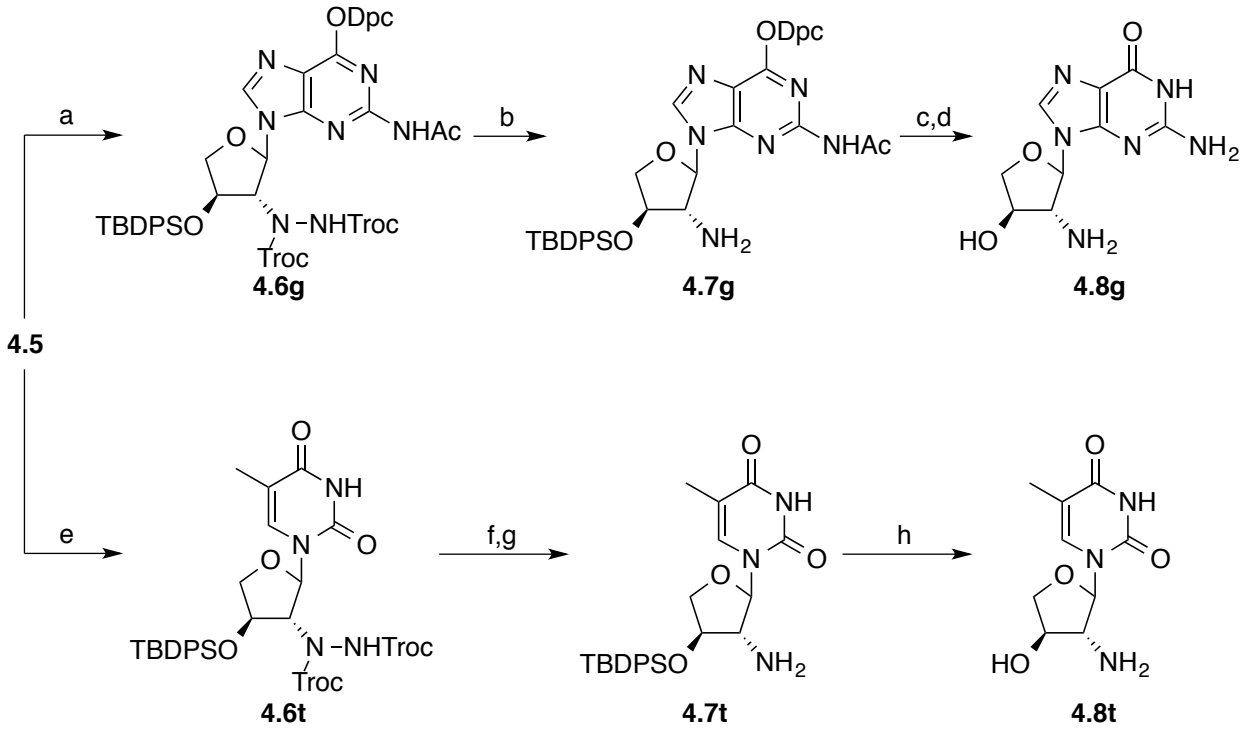


[a]Reagents and conditions: (a) *tert*-butyldiphenylchlorosilane, imidazole, DMF, room temperature, 23 h, 95%; (b) diisobutylaluminum hydride, THF, -78°C, 1 h, 95%; (c) methanesulfonyl chloride, triethylamine, dichloromethane, -40°C, 1 h; (d) reflux, 10 h, 74%; (e) bis(2,2,2-trichloroethyl) azodicarboxylate, 350 nm, cyclohexane, room temperature, 16 h, 83%.

Conveniently, the cycloaddition product **4.5** proved to be an excellent electrophile for one-pot nucleosidation catalyzed by trimethylsilyl trifluoromethanesulfonate (TMSOTf) (Wright and Dudycz 1984) (Scheme 4.2). Both the thymine and guanine nucleobases were glycosylated with high selectivity for the desired α anomer, likely indicating a mechanism involving direct displacement at C1. For guanine, we used the O^6-diphenylcarbamoyl protected nucleobase for high N9 regioselectivity (Robins et al. 1996). For the thymine hydrazide **4.6t**, the reported reduction conditions (Leblanc and Fitzsimmons 1989) with zinc powder in acetic acid with one equivalent of acetone were effective in producing the protected nucleoside **4.7t**. However, with the guanine hydrazide **4.6g** these conditions gave product degradation and reduction of the N2-acetyl group to an ethyl group. Therefore, we

screened various alternatives and ultimately found that a mixture of indium powder and aluminum foil in a mixture of ethanol and saturated aqueous ammonium chloride at reflux (Valluri et al. 2001; Mineno, Choi, and Avery 2002; Cicchi et al. 2003) gave the highest yield of protected nucleoside **4.7g**. Finally, we fully deprotected the intermediates to the free amino nucleosides **4.8g** and **4.8t**. Overall, the yield of **4.8g** and **4.8t** from lactone **4.1** was 12% and 24%, respectively, in 7 steps. The structure of each was solved by single crystal X-ray diffraction (Figure 4.2). The pseudorotation phase angle (Altona and Sundaralingam 1972) for nucleoside **4.8g** and the two molecules of **4.8t** in its crystal structure's asymmetric unit ranged from 0 to 23°, indicating a C3′-*endo* sugar pucker characteristic of residues in A-form DNA and RNA.

Scheme 4.2. Synthesis of 2′-amino-2′-deoxythreose nucleosides.[a]



Scheme 4.2 (continued). [a]Reagents and conditions: (a) N^2-acetyl-O^6-(diphenylcarbamoyl)guanine, bis(trimethylsilyl)acetamide, acetonitrile, 80°C, 10 min, cooled to 0°C and added **4.5** and then trimethylsilyl trifluoromethanesulfonate, 0°C, 15 min, room temperature, 1.5 h, 79%; (b) indium powder, aluminum foil, 2:1 EtOH-sat. aq. ammonium chloride, reflux, 3 h, 45%; (c) tetrabutylammonium fluoride, THF, room temperature, 3 h; (d) ammonium hydroxide, 35°C, 22 h, 62%; (e) **4.5**, thymine, bis(trimethylsilyl)acetamide, acetonitrile, 80°C, 10 min, cooled to 0°C and added trimethylsilyl trifluoromethanesulfonate, 0°C, 30 min, room temperature, 45 min, 91%; (f) zinc powder, acetic acid, room temperature, 1 h; (g) acetone, zinc powder, 5 h, 71%; (h) tetrabutylammonium fluoride, THF, room temperature, 3 h, 68%.

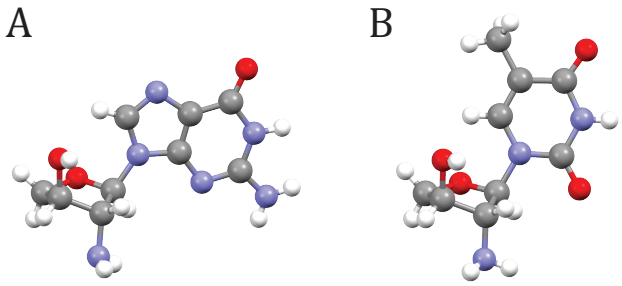


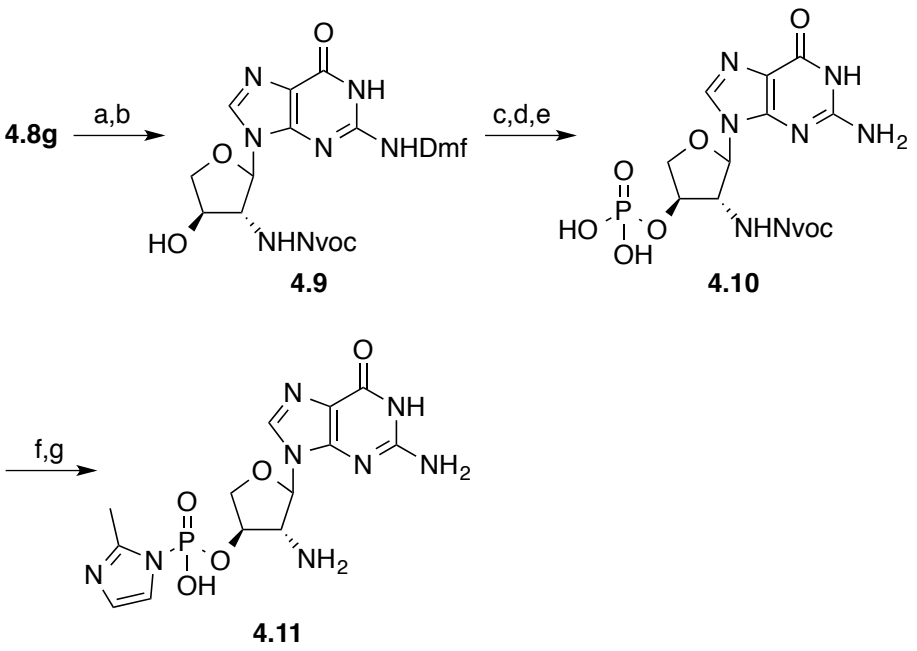
Figure 4.2. X-Ray crystal structures of the 2′-amino-2′-deoxythreose nucleosides with an (A) guanine (**4.8g**) and (B) thymine (**4.8t**) nucleobase. Both furanose rings have a C3′-*endo* sugar pucker.

Synthesis of 2-MeImpntG

Previously, we have synthesized activated amino nucleotides using phosphorus(V) oxychloride without protecting the nucleobase (Schrum et al. 2009; Zhang et al. 2013).

However, since the 3′-hydroxyl group of nucleosides **4.8g** and **4.8t** is a secondary hydroxyl and it is further sterically hindered by its proximity to the nucleobase, this reaction was sluggish and preferentially phosphorylated the nucleobase exocyclic amine. Therefore, we developed a new route to the 3′-phosphor-2-methylimidazolide activated nucleotide of **4.8g** (Scheme 4.3). First, we protected the exocyclic amine of the guanine moiety with the base labile dimethylformamidine (Dmf) group and the 2′-amine with the photolabile 4,5-dimethoxy-2-nitrobenzyloxycarbonyl (Nvoc) group to afford protected nucleoside **4.9**. To phosphorylate the 3′-hydroxyl group we used a phosphitylation-oxidation scheme since the 3′-hydroxyl group of TNA nucleosides has proven reactive toward phosphitylation reagents (Ferencic et al. 2004). Finally, we activated the phosphate of nucleotide **4.10** using a standard procedure developed by Orgel and colleagues (Joyce, Inoue, and Orgel 1984) and deprotected the 2′-amine with 350 nm irradiation. We used a pH of 10.5 during deprotection to minimize the hydrolysis rate (Kanavarioti et al. 1989). At 3%, the overall yield of 2-MeImpntG (**4.11**) from nucleoside **4.8g** was poor, but was sufficient for further characterization and primer extension experiments. We measured the rate of hydrolysis of **4.11** at pH 7.5 and 20^{o}C by ^{31}P NMR spectroscopy and observed a pseudo-first order rate constant of 4.47±0.05 x10^{-3} h^{-1} (estimate ± standard error), corresponding to a half-life of 6.47±0.07 days (Figure 4.3).

Scheme 4.3. Synthesis of the activated nucleotide 2-MeImpntG.[a]



[a]Reagents and conditions: (a) 4,5-dimethoxy-2-nitrobenzyl chloroformate, *N*,*N*-diisopropylethylamine, DMF, room temperature, 3.5 h; (b) *N*,*N*-dimethylformamide dimethyl acetal, DMF, room temperature, 6 h, 50%; (c) bis(2-cyanoethyl)-*N*,*N*-diisopropyl phosphoramidite, 5-(ethylthio)tetrazole, *N*,*N*-diisopropylethylamine, acetonitrile, room temperature, 2.7 h; (d) *tert*-butyl hydroperoxide, room temperature, 1.3 h; (e) ammonium hydroxide, 55°C, 18 h; (f) 2-methylimidazole, triethylamine, triphenylphosphine, 2,2′-dipyridyl disulfide, 3:2 DMSO-DMF, room temperature, 3 h; (g) 350 nm irradiation, water pH 10.5, 4°C, 16.5 h, 6% from **4.9**.

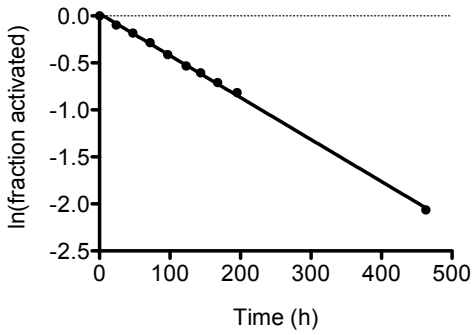


Figure 4.3. Hydrolysis of the activated nucleotide 2-MeImpntG (**4.11**) measured by ^{31}P NMR at pH 7.5 and 20°C. The natural logarithm of the fraction of activated nucleotide remaining was fit to a line to get a pseudo-first order rate constant of 4.47±0.05 $x10^{-3}$ h^{-1} (estimate ± standard error).

Primer extension of 2-MeImpntG on RNA, DNA and TNA templates

Once we had synthesized the activated 2′-NH_2-TNA nucleotide **4.11** we tested its reactivity in non-enzymatic primer extension experiments. To allow for a direct comparison to previous results from other amino-modified nucleic acid systems (Schrum et al. 2009; Zhang et al. 2013) we used the same reaction conditions: 5 mM monomer, pH 7.5, 150 mM NaCl and 100 mM *N*-hydroxyethylimidazole (HEI) as a catalyst. Similarly, we used a DNA primer with a 3′-amino terminus and 3′-C_4A_2-5′ RNA, DNA and TNA templates (Figure 4.4A). Activated nucleotide **4.11** was an effective monomer for the non-enzymatic primer extension reaction, with only 28% of the primer remaining after 1 h and 42% of the primer converted to full-length +4 product after 6 h on the RNA template (Figure 4.4B).

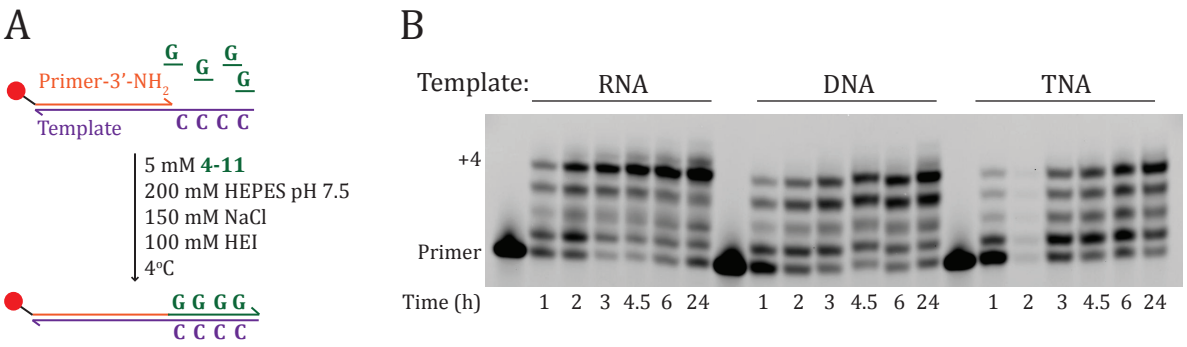


Figure 4.4. Primer extension reaction with the activated nucleotide 2-MeImpntG (**4.11**) on RNA, DNA and TNA templates. (A) Scheme of the reaction conditions. (B) PAGE gel analysis of the primer extension products.

To compare the reaction rates on the different templates we first fit the fraction of primer remaining to a first order exponential decay (Figure 4.5). This analysis gave pseudo-first order rate constants of 1.5±0.1, 0.97±0.06, and 0.57±0.06 h^{-1} for the RNA, DNA, and TNA templates, respectively. As expected, the RNA template, which binds more tightly to TNA (Schoning et al. 2000), was a more effective template than DNA. However, the TNA template, which we expected to form a pre-organized single strand for favorable primer extension, was actually the least effective template. The reactions reached a plateau with 8 to 9% primer remaining, possibly indicating a population of primer that was chemically modified or not properly bound to the template.

It was clear from the gel analysis that the different steps of primer extension occurred at different rates. Therefore, we modeled each step as a pseudo-first order reaction and found the rate constants that gave the best agreement to the observed values after simulating the kinetic model (Table 4.1). The estimated rate constants from this method were in agreement with those found from fitting an exponential decay and the

general trend of RNA>DNA>TNA held for all steps, except for the production of +4 product, which was faster on the TNA template than on DNA. Consistent with the low concentration of the +2 extension product, the rate constant for formation of the +2 product was less than that for its consumption on all templates, leading to a steady state concentration. Indeed, the formation of the +3 product was the fastest step on each template.

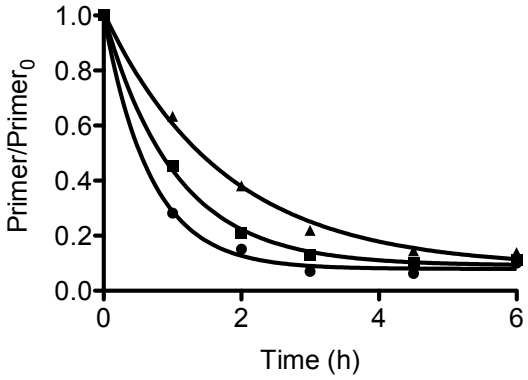


Figure 4.5. Consumption of primer from the non-enzymatic polymerization of 5 mM 2-MeImpntG (**4.11**) on a C_4 template at pH 7.5 and 4°C with 100 mM HEI catalyst. The fraction of primer remaining (Primer/Primer$_0$) was plotted against time and fit to a single exponential decay to give pseudo-first order rate constants of 1.5±0.1, 0.97±0.06 and 0.57±0.06 h^{-1} (estimate ± standard error) and plateau values of 8±1, 9±1 and 9±3% for the RNA (•), DNA (■) and TNA (▲) templates, respectively.

Table 4.1. Estimated pseudo-first order rate constants for the primer extension reaction of 2-MeImpntG (**4.11**) on C_4 templates as determined by differential equation modeling.

	k_{obs} (h^{-1})		
Step	RNA	DNA	TNA
Primer → +1	1.52	0.96	0.57
+1 → +2	0.96	0.75	0.40
+2 → +3	1.55	1.14	0.67
+3 → +4	0.60	0.32	0.46
+4 → +5	0.12	0.14	0.06

Discussion

We have developed a novel synthesis of 2′-NH_2-TNA nucleosides that allowed us to test the reactivity of the 2-MeImpntG monomer (**4.11**) in non-enzymatic primer extension reactions on RNA, DNA, and TNA templates. The synthesis was based on a [4+2] cycloaddition reaction between a glycal and azodicarboxylate that was previously developed to synthesize 2-amino sugars (Leblanc and Fitzsimmons 1989). We found that the reaction worked well for the four-carbon glycal **4.4** if we omitted the dichloromethane co-solvent from the published procedure, giving a yield of 83% of the desired regio- and stereoisomer (Scheme 4.1). The synthetic route to the nucleoside was simplified greatly by the use of cycloadduct **4.5** directly as a nucleosidation substrate to give the 2′-hydrazide nucleoside series **4.6** with high anomeric selectivity (Scheme 4.2). Although the standard hydrazide reduction conditions worked well for the thymine nucleoside **4.6t**, it was not compatible with the nucleobase protecting groups for the guanine analog **4.6g**. Therefore, to access the 2′-amino nucleoside **4.8g** we had to use milder and more selective reducing conditions with indium metal as a single electron donor.

Overall, **4.8g** and **4.8t** were synthesized in 12% and 24% yield, respectively, from lactone **4.1** in 7 steps. For comparison, Eschenmoser and colleagues synthesized **4.8t** in 9 steps from L-ascorbic acid in 10% overall yield (Schoning et al. 2000; Schoning et al. 2002; Ferencic et al. 2004). Both nucleosides crystallized with a C3′-*endo* sugar pucker characteristic of residues in A-form DNA and RNA (Figure 4.2), whereas TNA residues prefer a C4′-*exo* sugar pucker in the context of a duplex (Wilds et al. 2002; Pallan et al. 2003; Ebert et al. 2008). The C3′-*endo* sugar pucker, as well as other sugar puckers, has also been seen in other crystallized TNA nucleosides (Schoning et al. 2000), suggesting that the C4′-*exo* sugar pucker is preferred only in the context of a duplex. Since both the C3′-*endo* and C4′-*exo* sugar puckers hold the 2′- and 3′-substituents in the extended axial conformation required for the duplex, the C4′-*exo* sugar pucker is likely required for the optimal orientation of the nucleobase and phosphodiester backbone.

The cycloaddition-nucleosidation-reduction scheme we have developed could potentially be applied to other 2′-amino nucleosides. The stereoselectivity of the cycloaddition reaction can be maintained with 3-deoxy glycals by using a bulky protecting group on the 4- and 5-hydroxyl group (Leblanc et al. 1989). Therefore, it could potentially be used to access the 2′-amino-2′,3′-dideoxyribose nucleoside series, which performs well in non-enzymatic polymerization experiments, but is currently difficult to synthesize (Schrum et al. 2009). The more common 2′-hydroxyl nucleosides could be accessed by using phenanthrenequinone instead of azodicarboxylates. It has previously been shown that phenanthrenequinone can undergo similar stereoselective [4+2] cycloaddition chemistry with glycals to form new C1-O and C2-O bonds (Lichtenthaler, Weimer, and

Immel 2004), although it is not known if the cycloadduct formed is a good substrate for nucleosidation.

Next, we turned our attention to synthesizing the activated monomer 2-MeImpntG (**4.11**) and testing its reactivity in non-enzymatic primer extension experiments. Unfortunately, since the 3′-hydroxyl group of **4.8g** is secondary and sterically hindered due to its proximity to the nucleobase, it was not sufficiently reactive in our standard phosphorus(V) oxychloride phosphorylation-activate conditions (Schrum et al. 2009; Zhang et al. 2013). Therefore, we protected the 2′-amine and nucleobase and used a phosphitylation-oxidation-activation route instead (Scheme 4.3). At 3%, the overall yield from **4.8g** was low, but was sufficient for our purposes. It should be possible to improve the yield by using an alternate activation method, for example, 1,1′-carbonyldi-2-methylimidazole. We confirmed that the activated nucleotide **4.11** is not able to cyclize by following its decomposition by ^{31}P NMR (Figure 4.3). Indeed, it hydrolyzed with a half-life of 6.47±0.07 days at pH 7.5 and 20^{o}C, as compared to 5.3±0.2 days for guanosine 5′-phosphor-2-methylimidazolide (Chapter 3). The slower hydrolysis rate could be due to the guanosine nucleobase blocking the attack of water on the 3′-phosphate of **4.11**.

Finally, we demonstrated that monomer **4.11** could extend from a primer on C_4 RNA, DNA, and TNA templates (Figure 4.4). This finding is the first demonstration of non-enzymatic polymerization by a threose-based nucleotide and is critical for its potential use as a self-replicating genetic material. The pseudo-first order rate constants for the first addition to the primer ranged from 0.57±0.06 to 1.5±0.1 h^{-1} (Figure 4.5). For comparison, in the 2′-amino and 3′-amino-2′,3′-dideoxyribose systems the activated guanine nucleotides completely copy a C_4 RNA template in under 5 minutes, indicating a rate

constant of at least 42 h^{-1}. It is possible that the constrained sugar of **4.11** limits its ability to adopt the geometry necessary for the phosphoryl transfer when bound to the template. We also found that the reaction rate was greatest for the third nucleotide addition step even though there were only two template positions available for cooperative base stacking (Table 4.1). This trend was also observed for the 3′-amino-2′,3′-dideoxyribose system (Zhang et al. 2013) and may be due to the change in helical geometry as the amino nucleotides extend from the DNA primer.

Interestingly, the ranking of the polymerization rates on the different templates was RNA>DNA>TNA. We expected polymerization to be faster on the A-form RNA template than the B-form DNA template since TNA prefers an A-form geometry due to its shortened backbone (Pallan et al. 2003). However, we similarly expected polymerization to be fastest on the TNA template. TNA is also a worse template than RNA for the polymerization of activated ribonucleotides (Heuberger and Switzer 2006). The restrained geometry of TNA either makes it less able to bind tightly to nucleotides as a template or for the TNA nucleotides to adopt the geometry necessary for phosphoryl transfer once bound. These alternate hypotheses could be tested by measuring the rate of polymerization at a range of monomer concentrations on the different templates to differentiate monomer affinity from maximal rate. Combined, our results show that 2′-NH_2-TNA could potentially be used as a self-replicating genetic material, but its relatively slow rate of polymerization, especially on the TNA template, makes other systems more attractive. If these results hold for TNA monomers it would suggest that TNA would be inferior to RNA as a genetic polymer in the origin of life, which may be why it was ultimately not used despite the likely co-production

of threose- and ribose-based nucleosides through prebiotic chemistry (Kim et al. 2011; Islam et al. 2013).

To further test the potential of 2′-NH_2-TNA to function as a self-replicating genetic polymer it will be necessary to synthesize and test the activated cytosine, adenine and thymine monomers to ensure that all four bases can be copied. The fidelity of copying is also important for the accurate replication of genetic information, and the relative rigidity of the 2′-NH_2-TNA backbone could make it particularly selective for the complementary nucleotide. 2′-NH_2-TNA could also have different properties as a template than TNA as it forms weaker duplexes than RNA, DNA and TNA (Ferencic et al. 2004). A weaker duplex is preferable as it facilitates strand separation for multiple rounds of replication, but it could also correspond to weaker monomer binding.

Methods and Materials

General

All reagents and solvents were purchased from Sigma-Aldrich, Alfa Aesar or Toronto Research Chemicals, except for deuterated solvents, which were purchased from Cambridge Isotope Laboratories. Flash chromatography was performed on a Biotage SP1 instrument with HP-Sil columns. NMR spectroscopy was performed on a 400 MHz Varian spectrometer (Oxford AS-400) operating at 25^{o}C, unless otherwise specified. Spectra were referenced to the solvent peak according to published values (Gottlieb, Kotlyar, and Nudelman 1997), except for ^{31}P spectra, which were referenced to orthophosphate or phosphoric acid (0 ppm). High-resolution mass spectrometry was carried out on a Waters Q-TOF micro LC-MS or an Agilent 6520 QTOF LC-MS. Low resolution ESI-MS was

performed on a Bruker Esquire 6000 with direct injection. Preparative HPLC purification was performed on a Varian ProStar instrument with a Varian Dynamax Microsorb C18 column (250x21.4 mm).

*X-ray crystallography of **4.8t***

An analytical sample of **4.8t** was crystallized from water with ethanol vapor diffusion. Data was collected from a crystal mounted on a diffractometer at 100 K. The intensities of the reflections were collected by means of a Bruker APEX II CCD along with the D8 Diffractometer (30 KeV, λ = 0.413280 Å), and equipped with an Oxford Cryosystems nitrogen open flow apparatus. The collection method involved 0.5° scans in Phi at -5° in 2Θ. Data integration down to 0.82 Å resolution was carried out using SAINT V7.46 A (Bruker diffractometer, 2009) with reflection spot size optimization. Absorption corrections were made with the program SADABS (Bruker diffractometer, 2009). The structure was solved by the direct methods procedure refined by least-squares methods again F^2 SHELXS-97 and SHELXL-97 (Sheldrick, 2008). Non-hydrogen atoms were refined anisotropically, and hydrogen atoms were allowed to ride on the respective atoms.

*X-ray crystallography of **4.8g***

An analytical sample of **4.8g** was crystallized from water with ethanol vapor diffusion. Data was collected from a crystal mounted on a diffractometer at 100 K. The intensities of the reflections were collected by means of a Bruker APEX II DUO CCD diffractometer (Cu_K radiation, λ = 1.54178 Å), and equipped with an Oxford Cryosystems

nitrogen flow apparatus. The collection involved 1.0° scans in ω at 30, 55, 80 and 105° in 2θ. Data analysis was carried out as per **4.8t**.

Primer extension reaction

The conditions for the primer extension reaction were as follows: 200 mM HEPES pH 7.5, 150 mM NaCl, 100 mM *N*-hydroxylethylimidazole (HEI) pH 7.5, 5 mM 2-MeImpntG (**4.11**), 0.3 μM primer, 1.5 μM template, 4°C. The primer sequence was 5′-TAMRA-GCG TAG ACT GAC TGG-3′-NH_2, the RNA template sequence was 5′-AA**C CCC** CCA GUC AGU CUA CGC-3′ and the DNA template sequence was 5′-AA**C CCC** CCA GTC AGT CTA CGC-3′, where the bold sequence is the single-stranded overhang. The TNA template was DNA, except for the underlined region, which was TNA: 5′-AA**C CCC** CCA GTC AGT CTA CGC-3′. The primer was generously provided by Dr. Sergei Gryaznov (Geron Corp.), the TNA template was generously provided by Dr. John Chaput (Arizona State University) and the RNA and DNA templates were purchased from IDT DNA. The primer was purified by PAGE, reverse phase HPLC and ion exchange HPLC prior to use.

The total reaction volume was 30 μL. At each time point a 5 μL aliquot was taken, mixed with 10 μL kill buffer (100 mM EDTA, 4.8 M urea, 1X TBE) and ethanol precipitated. Time points were taken at 1, 2, 3, 4.5, 6 and 24 h. Samples were analyzed by 20% urea-TBE denaturing PAGE and imaged on a Typhoon Scanner 9400 (GE Healthcare) and integrated using ImageQuant TL software (GE Healthcare). The fraction of the total integration of each lane represented by unreacted primer was plotted against time (omitting the 24 h time point and including a zero time point) and fit to a single exponential decay to get a pseudo-first order rate constant and plateau using Prism software (GraphPad).

Estimating primer extension rate constants by differential equation modeling

To estimate the rate constant for each monomer addition step in the primer extension reactions we built a simple model based on the following set of differential equations, where k_n is the pseudo-first order rate constant for the conversion of primer+n to primer+n+1:

$$\frac{d[Primer]}{dt} = -k_0 \times [Primer]$$
$$\frac{d[Primer+1]}{dt} = k_0 \times [Primer] - k_1 \times [Primer+1]$$
$$\frac{d[Primer+2]}{dt} = k_1 \times [Primer+1] - k_2 \times [Primer+2]$$
$$\frac{d[Primer+3]}{dt} = k_2 \times [Primer+2] - k_3 \times [Primer+3]$$
$$\frac{d[Primer+4]}{dt} = k_3 \times [Primer+3] - k_4 \times [Primer+4]$$
$$\frac{d[Primer+5]}{dt} = k_4 \times [Primer+4]$$

We then developed a MATLAB function, termed "all band analysis" (ABA), that takes a matrix of gel integrations of a primer extension reaction (columns are time points with primer in bottom row) and a vector of the times represented by each column. For each step in the primer extension reaction it scans a range of k values and selects the value that has the least squares error between the predicted and observed integrations. It then fixes the k value for that step and repeats the process for the next step. Since about 9% of the primer appeared to be unreactive (Figure 4.5) we subtracted this amount from each primer

integration before modeling. The code for this script is given below, and the most up-to-date version is available at https://github.com/jcraigblain:

```
function kays = ABA(title, k0, kstep, kf, time_points, integrations)
%Fit k values to multiple steps in a primer extension reaction
%Takes a matrix of gel integrations (increasing time left to right,
%increasing extension length bottom to top)
%Fits a k value to each extension step one at a time

%normalize gel lanes
column_sums=sum(integrations,1);
for i=1:size(integrations,1)
   integrations(:,i)=integrations(:,i)/column_sums(i);
end

times = time_points - time_points(1); %t0 now 0

flipped_integrations = flipud(integrations); %primer now at top

%for each product length n, add all integrations for >=n+1 into n+1 band
%then fit a kay value for n->n+1
kays=[];
[num_bands,num_timepoints] = size(flipped_integrations);
for b=2:num_bands
    reduced_integrations = [flipped_integrations(1:b-1,:);
sum(flipped_integrations(b:end,:),1)];
    kays(b-1) = ABA_kScan(k0, kstep, kf, times, reduced_integrations, kays);
end

%Model fitted k values
initial_integrations = reduced_integrations(:,1);
model_output = initial_integrations;
for t=2:length(times)
    model_result = ABA_RateODE(kays,times(t),initial_integrations);
    model_output(:,t) = model_result(end,:)';
end

%Plot model output
model_output = flipud(model_output);
model_heatmap_data = flipud(model_output);
y_labels = 0:length(kays);
hm1 =
HeatMap(model_heatmap_data,'RowLabels',y_labels,'ColumnLabels',time_points,'S
ymmetric',false,'DisplayRange',1,'Colormap',colormap(flipud(gray)),'Standardi
ze','none');
addXLabel(hm1,'Time (h)');
addYLabel(hm1,'Extension');
addTitle(hm1,strcat(title,' Model'));
plot(hm1,figure(1))

%Plot input data
input_heatmap_data = flipud(integrations);
y_labels = 0:length(kays);
hm2 =
HeatMap(input_heatmap_data,'RowLabels',y_labels,'ColumnLabels',time_points,'S
```

```
ymmetric',false,'DisplayRange',1,'Colormap',colormap(flipud(gray)),'Standardi
ze','none');
addXLabel(hm2,'Time (h)');
addYLabel(hm2,'Extension');
addTitle(hm2,strcat(title,' Data'));
plot(hm2,figure(2))
end

function kmin = ABA_kScan(k0, kstep, kf, time_points, integrations,kays)
%scans a range of k values for a particular step in the primer extension
%reaction and selections the one that gives the least squared error
%from the measured integrations

k_scan = k0:kstep:kf;
r = zeros(length(k_scan));

initial_integrations = integrations(:,1);
for i=1:length(k_scan)
    test_k=k_scan(i);
    for j=2:length(time_points)
        model = ABA_RateODE([kays test_k], time_points(j),
initial_integrations);
        r(i) = r(i) + sum((integrations(:,j)-model(end,:)').^2);
    end
end

[rminn, index] = min(r);
kmin = k_scan(index(1));

end

function x = ABA_RateODE(kays,tf,initial_integrations)
%Kinetic model of primer extension reaction
%Models each step as a pseudo-first order reaction
%Works for any n

%Invoke ODE Solver

[t, x] = ode15s(@derivatives, [0 tf], initial_integrations);

%Defining ODE system for ode solver
function xdot = derivatives(t,x)
    num_bands = length(kays)+1;
    xdot = zeros(num_bands,1);
    xdot(1) = -kays(1)*x(1); %consumption of primer
    for i=2:num_bands-1
       xdot(i) = kays(i-1)*x(i-1)-kays(i)*x(i); %production and consumption
of intermediate bands
    end
    xdot(num_bands) = kays(end)*x(num_bands-1); %production of last product
end

end
```

Kinetics of 2-MeImpntG hydrolysis

The rate of hydrolysis of 2-MeImpntG (**4.11**) was measured under the following conditions: 3 mM **4.11**, 100 mM HEPES pH 7.5, 5 mM potassium phosphate pH 7.5, 10% D_2O, 20°C. Time points were taken daily for 8 days and then once more after 19 days. At each time point a ^{31}P NMR spectrum was measured with 256 scans. The phosphate signal was set to a chemical shift of 0 ppm and an integration of 1. The integrations of **4.11** (-12.02 ppm) and the hydrolyzed product (1.15 ppm) were then measured. The natural logarithm of the fraction of the sum of the two integrations that was **4.11** was plotted against time and fit to a line using Prism software (GraphPad). The slope of the line was taken as the pseudo-first order rate constant for hydrolysis.

Synthetic procedures

(*R*)-4-((*tert*-butyldiphenylsilyl)oxy)dihydrofuran-2(3*H*)-one (4.2). A mixture of lactone **4.1** (10 g, 94.0 mmol) and imidazole (16.7 g, 245 mmol) in DMF (36 mL) under argon was cooled on ice and then TBDPSCl (32 mL, 119 mmol) was added slowly. The mixture was warmed to 20°C, stirred for 23 h and then poured into ddH_2O (400 mL). The product was extracted with CH_2Cl_2 (3 x 250 mL) and the combined organic fractions were washed with brine (200 mL), dried over $MgSO_4$, filtered and concentrated to 100 mL. The concentrate was filtered through a bed of silica with CH_2Cl_2 and then triturated with hexanes to afford protected lactone **4.2** (28.3 g, 95%) as a white solid: R_f =0.69, CH_2Cl_2; ^1H NMR (400 MHz, $CDCl_3$) δ 7.62 (d, *J* = 6.9 Hz, 4H, H-Ph), 7.46 (dd, *J* = 7.2 Hz, 2H, H-Ph), 7.40 (dd, *J* = 7.1 Hz, H-Ph), 4.56 (m, 1H, H-C3), 4.21 (dd, 1H, *J* = 9.8, 2.5 Hz, H-C4), 4.16 (dd, 1H, *J* = 9.8, 4.6 Hz, H-C4), 2.50 (d, 2H, *J* = 3.9 Hz, H-C2), 1.06 (s, 9H, H-*t*-Bu); ^{13}C NMR (100 MHz,

$CDCl_3$) δ 175.9, 135.8, 133.0, 130.4, 128.2, 75.9, 69.2, 38.1, 26.9, 19.2; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{20}H_{24}O_3Si$, 341.1568, obs. 341.1563.

(4*R*)-4-((*tert*-butyldiphenylsilyl)oxy)tetrahydrofuran-2-ol (4.3). Protected lactone **4.2** (10.02 g, 29.4 mmol) was dissolved in anhydrous THF (100 mL) under argon and cooled to -78°C with an acetone-dry ice bath. A 1 M solution of DIBAL-H in toluene (38 mL, 38 mmol) was added dropwise over 1 h. The reaction was quenched with methanol (3 mL), warmed to room temperature. Saturated sodium potassium tartrate (150 mL) was added and the mixture was poured into 400 mL EtOAc, stirred for 2 h. The organic layer was washed with water (150 mL), brine (150 mL), dried over $MgSO_4$ and filtered and then the crude was purified by flash chromatography (5 - 20% EtOAc in hexanes) to afford protected lactol **4.3** (9.58 g, 95%) as a clear oil: R_f = 0.28, 1:4 EtOAc-hexanes; ^{1}H NMR (400 MHz, DMSO-d_6) anomer 1: δ 7.63 - 7.54 (m, 4H, H-Ph), 7.48-7.37 (m, 6H, H-Ph), 6.06 (d, *J* = 5.1 Hz , 1H, H-O-C1), 5.44 (m, 1H, H-C1), 4.49 (m, 1H, H-C3), 3.76 (dd, *J* = 9.1, 4.9 Hz, 1H, H-C4), 3.55 (dd, *J* = 9.2, 2.4 Hz, 1H, H-C4), 1.94 (m, 1H, H-C2), 1.77 (ddd, *J* = 13.3, 6.3, 3.0 Hz, 1H, H-C2), 0.97 (s, 9H, H-*t*-Bu); anomer 2: δ 7.63 - 7.54 (m, 4H, H-Ph), 7.48-7.37 (m, 6H, H-Ph), 6.11 (d, *J* = 5.0, 1H, H-O-C1), 5.20 (m, 1H, H-C1), 4.28 (m, 1H, H-C3), 3.68-3.60 (m, 2H, H-C4), 2.10 (m, 1H, H-C2), 1.67 (m, 1H, H-C2), 0.98 (s, 9H, H-*t*-Bu); ^{13}C NMR (100 MHz, DMSO-d_6) δ 135.2, 135.1, 133.3, 133.2, 129.9, 127.9, 97.6, 97.3, 72.9, 72.6, 72.1, 71.5, 43.3, 42.4, 26.7, 18.7, 18.6; HRMS (m/z): $[M+Na^+]^+$ calc. for $C_{20}H_{26}O_3Si$, 365.1543, obs. 365.1544.

(*R*)-*tert*-butyl((2,3-dihydrofuran-3-yl)oxy)diphenylsilane (4.4). Protected lactol **4.3** (8.94 g, 27.6 mmol) and Et_3N (14.6 mL, 104.7 mmol) were dissolved in anhydrous CH_2Cl_2 (200 mL) under argon then cooled to -40°C with an acetonitrile-dry ice bath. MsCl (2.7 mL, 34.9 mmol) was added slowly and the reaction was stirred at -40°C for 1 h, then

warmed to room temperature and refluxed for 10 h. The reaction mixture was concentrated to 100 mL and then purified by flash chromatography (1:1 CH_2Cl_2-hexanes) to afford protected glycal **4.4** (6.59 g, 74%) as a clear oil: R_f = 0.57, 1:1 CH_2Cl_2-hexanes; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.68-7.64 (m, 4H, H-Ph), 7.45-7.41 (m, 2H, H-Ph), 7.40-7.36 (m, 4H, H-Ph), 6.49 (apt. d, *J* = 2.6 Hz, 1H, H-C1), 5.01 (dddd, *J* = 7.2, 2.6, 2.6, 0.9 Hz, 1H, H-C3), 4.96 (dd, *J* = 2.6, 2.6 Hz, 1H, H-C2), 4.23 (dd, *J* = 10.5, 2.6 Hz, 1H, H-C4), 4.04 (dd, *J* = 10.5, 7.20 Hz, 1H, H-C4), 1.04 (s, 9H, H-*t*-Bu); ^{13}C NMR (100 MHz, DMSO-d_6) δ 150.5, 135.5, 135.4, 133.8, 133.6, 130.2, 130.1, 128.3, 128.2, 103.7, 77.1, 74.8, 27.0, 19.0; HRMS (m/z): $[M+Na^+]^+$ calc. for $C_{20}H_{24}O_2Si$, 347.1, obs. 347.1.

(4a*R*,7*R*,7a*R*)-2,2,2-trichloroethyl 7-((*tert*-butyldiphenylsilyl)oxy)-3-(2,2,2-trichloroethoxy)-4a,6,7,7a-tetrahydro-1*H*-furo[3,2-*e*][1,3,4]oxadiazine-1-carboxylate (4.5). Protected glycal **4.4** (1.00 g, 9.25 mmol) and bis(2,2,2-trichloroethyl)azodicarboxylate (1.65 g, 13.0 mmol) were dissolved in 18 mL anhydrous cyclohexane under argon in a sealed 40 mL glass vial. The solution was irradiated at 350 nm while stirred at r.t. for 16 h. Volatiles were removed *in vacuo* and the crude residue was purified by flash chromatography (0-10% EtOAc in hexanes) to afford cycloadduct **4.5** (1.83 g, 83%) as a colorless foam: R_f = 0.52, 3:7 ether-hexanes; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.6-7.3 (m, 10H, H-Ph), 5.69 (d, *J* = 4.6 Hz, 1H, H-C1), 5.20 (dd, *J* = 4.6, 6.6 Hz, 1H, H-C2), 4.70 (br. s, 4H, H_2C), 4.43 (apt. q, *J* = 6.4 Hz, 1H, H-C3), 3.90 (dd, *J* = 9.3, 7.0 Hz, 1H, H-C4), 3.80 (dd, *J* = 9.4, 5.9 Hz, 1H, H-C4), 1.07 (s, 9H, H_3C); ^{13}C NMR (100 MHz, $CDCl_3$) δ 135.8, 135.8, 135.3, 134.9, 132.2, 130.4, 130.3, 129.8, 128.1, 128.1, 127.9, 99.3, 95.2, 94.2, 74.2, 26.9, 26.7, 19.2; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{26}H_{28}Cl_6N_2O_6Si$, 702.9921, obs. 702.9895.

Bis(2,2,2-trichloroethyl) 1-((3*R*,4*R*)-2-(2-acetamido-6-((diphenylcarbamoyl)-oxy)-9*H*-purin-9-yl)-4-((*tert*-butyldiphenylsilyl)oxy)tetrahydrofuran-3-yl)hydrazine-1,2-dicarboxylate (4.6g). N^2-Acetyl-O^6-(diphenylcarbamoyl)guanine (2.15 g, 5.54 mmol) was suspend in anhydrous acetonitrile (40 mL) under argon. Bis(trimethylsilyl)acetamide (2.8 mL, 11.5 mmol) was added and the suspension was stirred at 80°C for 10 min to dissolve the silylated guanine. The solution was cooled to 0°C and then cycloadduct **4.5** (3.26 g, 4.62 mmol) was added in acetonitrile (30 mL) followed by the dropwise addition of TMSOTf (1.08 mL, 5.98 mmol). The reaction was stirred at 0°C for 15 min and then at r.t. for 1.5 h. Volatiles were removed *in vacuo* and the residue was dissolved in EtOAc (100 mL), washed with 5% $NaHCO_3$ (2x100 mL) and brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to a crude. The crude was purified by flash chromatography (20 to 60% EtOAc in hexanes) to afford protected hydrazide **4.6g** as a white foam (4.01 g, 79%). R_f = 0.72, 6% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6, 40°C) Note: broad multiplets were observed due to stable rotamers δ 8.47 (br d, 1H, H-C8), 7.77-7.32 (m, 20H, H-Ph), 6.20 (s, 1H, H-C1′), 5.65 (br s, 1H, H-C3′), 5.00-4.30 (br m, 6H, H-C4′ and H-Troc), 3.87 (br m, 1H, H-C2′), 2.29 (br s, 3H, H-Ac), 1.04 (br s, 9H, H-*t*Bu); ^{13}C NMR (100 MHz, DMSO-d_6, 60°C) δ 169.9, 155.2, 153.8, 152.1, 149.7, 143.9, 143.5, 141.4, 135.5, 135.3, 135.0, 132.4, 132.1, 130.0, 129.9, 129.6, 129.5, 129.3, 128.8, 127.7, 127.3, 126.8, 126.4, 120.3, 95.1, 94.5, 74.9, 59.4, 54.5, 26.5, 26.3, 18.4; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{46}H_{44}Cl_6N_8O_9Si$, 1091.1204, obs. 1091.1254.

2-Acetamido-9-((3*R*,4*R*)-3-amino-4-((*tert*-butyldiphenylsilyl)oxy)tetrahydrofuran-2-yl)-9*H*-purin-6-yl diphenylcarbamate (4.7g). Protected hydrazide **4.6g** (4.54 g, 4.15 mmol), indium powder (1.90 g, 16.6 mmol) and aluminum foil (1.29 g, 47.9 mmol)

were suspended in EtOH (67 mL) and sat. aq. NH_4Cl (33 mL) and refluxed under argon with vigorous stirring for 3 h. The hot reaction mixture was then filtered through a short bed of celite, diluted with water (200 mL), adjusted to pH 11 with 1 M NaOH and then extracted with CH_2Cl_2 (3x100 mL). The combined organic phase was dried over $MgSO_4$, filtered and concentrated *in vacuo* to a crude. The crude was purified by flash chromatography (0 to 4% MeOH in CH_2Cl_2) to afford protected aminonucleoside **4.7g** as a white foam (1.35 g, 45%). R_f = 0.29, 6% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.56 (s, 1H, H-C8), 7.57-7.27 (m, 20H, H-Ph), 5.80 (d, *J* = 2.7 Hz, 1H, H-C1′), 4.17 (dd, *J* = 7.4, 3.3 Hz, 1H, H-C3′), 4.11 (dd, *J* = 9.1, 3.1 Hz, 1H, H-C4′), 4.04 (dd, *J* = 9.4, 5.1 Hz, 1H, H-C4′), 4.01 (m, 1H, H-C2′), 2.22 (s, 3H, H-Ac), 0.88 (s, 9H, H-*t*Bu); ^{13}C NMR (100 MHz, DMSO-d_6) δ 169.1, 155.1, 154.2, 152.1, 150.1, 143.5, 141.7, 135.3, 135.2, 132.6, 132.6, 130.1, 130.0, 128.0, 127.9, 120.4, 91.3, 78.4, 74.7, 63.6, 26.6, 24.8, 18.5; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{40}H_{41}N_7O_5Si$, 728.3011, obs. 728.2943.

9-(2′-Amino-2′-deoxy-α-L-threofuranosyl)guanine (4.8g). Protected aminonucleoside **4.7g** (234 mg, 321 mmol) was dissolved in THF (15 mL) under argon. A 1 M solution of TBAF in THF (0.48 mL, 480 mmol) was added and the reaction was stirred at room temperature for 3 h. Volatiles were removed *in vacuo* and the residue was transferred to a pressure vessel in MeOH (5 mL), 28% NH_4OH (15 mL) was added and the sealed flask was heated at 35°C for 22 h. Volatiles were removed *in vacuo* and the residue was purified by flash chromatography (2 to 20% MeOH in CH_2Cl_2) to afford aminonucleoside **4.8g** (50 mg, 62%) as a white solid. R_f = 0.04, 3:7 methanol-dichloromethane; ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.88 (s, 1H, H-C8), 5.54 (d, *J* = 3.4 Hz, 1H, H-C1′), 4.06 (dd, *J* = 8.8, 4.8 Hz, 1H, H-C4′), 4.00 (m, 1H, H-C3′), 3.89 (dd, *J* = 8.8, 3.2 Hz, 1H,

H-C4′), 3.55 (apt. t, *J* = 3.1 Hz, 1H, H-C2′); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.7, 154.5, 152.0, 137.0, 117.4, 90.5, 77.4, 75.0, 65.0; HRMS (m/z): $[M+H^+]^+$ calc. for $C_9H_{12}N_6O_3$, 253.1044, obs. 253.1045.

Bis(2,2,2-trichloroethyl) 1-((3*R*,4*R*)-4-((*tert*-butyldiphenylsilyl)oxy)-2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2*H*)-yl)tetrahydrofuran-3-yl)hydrazine-1,2-dicarboxylate (4.6t). Thymine (1.25 g, 9.91 mmol) and cycloadduct **4.5** were suspended in anhydrous acetonitrile (76 mL) under argon. Bis(trimethylsilyl)acetamide (7.25 mL, 29.7 mmol) was added and the suspension was stirred at 80°C for 10 min to dissolve the silylated thymine. The solution was cooled to 0°C and then TMSOTf (1.24 mL, 6.86 mmol) was added dropwise. The reaction was stirred at 0°C for 30 min and then at r.t. for 45 min. Volatiles were removed *in vacuo* and the residue was dissolved in CH_2Cl_2 (400 mL) and washed with 5% $NaHCO_3$ (300 mL). The aqueous phase was extracted further with CH_2Cl_2 (2x100 mL) and then dried over $MgSO_4$, filtered and concentrated *in vacuo* to a crude. The crude was purified by flash chromatography (0 to 2% MeOH in CH_2Cl_2) to afford protected hydrazide **4.6t** as a white foam (3.71 g, 91%). R_f = 0.40, 5% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6, 50°C) δ 11.30 (s, 1H, H-N3), 10.41 (br s, 1H, H-N-Troc), 7.63-7.35 (m, 11H, H-Ph and H-C6), 6.07 (br s, 1H, H-C1′), 5.00-4.65 (m, 6H, H_2-$CCCl_3$ and H_2-C4′), 4.07 (app d, *J* = 9.3 Hz, H-C3′), 3.76 (br s, 1H, H-C2′), 1.73 (s, 3H, H_2-CC5), 1.04 (s, 9H, H-*t*Bu); ^{13}C NMR (100 MHz, DMSO-d_6) δ 163.8, 155.0, 153.0, 150.2, 135.2, 132.4, 130.2, 128.0, 95.4, 95.0, 83.4, 75.2, 73.9, 70.8, 26.6, 18.6, 12.2; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{31}H_{34}Cl_6N_4O_8Si$, 829.0350, obs. 829.0341.

1-((3*R*,4*R*)-3-Amino-4-((*tert*-butyldiphenylsilyl)oxy)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1*H*,3*H*)-dione (4.7t). Protected hydrazide **4.6t** (250 mg, 301

μmol) was dissolved in glacial acetic acid (10 mL) under argon. Zinc powder (950 mg) was added and the suspension was stirred vigorously for 1 h at room temperature. Acetone (0.04 mL, 544 μmol) was added and the reaction was stirred for another 1 h. Finally, more zinc powder (750 mg) was added and the reaction was stirred for another 4 h. The suspension was filtered, washed with MeOH, concentrated *in vacuo* and coevaporated with toluene to obtain a crude foam. The crude was purified by flash chromatography (0 to 2% MeOH in CH_2Cl_2 with 0.1% Et_3N to afford protected nucleoside **4.7t** as a white foam (99 mg, 71%). R_f = 0.35, 6% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6, 40^{o}C) δ 7.62-7.42 (m, 11H, H-Ph and H-C6), 5.50 (d, *J* = 2.1 Hz, H-C1′), 4.10-4.00 (m, 3H, H-C4′ and H-C3′), 3.43 (m, 1H, H-C2′), 1.73 (d, *J* = 1.2, H_3C-C5), 0.98 (s, 9H, H-*t*Bu); ^{13}C NMR (100 MHz, DMSO-d_6) δ 163.9, 150.5, 136.4, 135.2, 132.6, 130.0, 128.0, 108.3, 92.6, 78.5, 75.4, 64.4, 26.5, 18.6, 12.3; HRMS (m/z): $[M+H^+]^+$ calc. for $C_{25}H_{31}N_3O_4Si$, 466.2, obs. 466.1.

1-(2′-Amino-2′-deoxy-α-L-threofuranosyl)thymine (4.8t). Protected nucleoside **4.7t** (203 mg, 450 μmol) was dissolved in anhydrous THF (15 mL) under argon. A 1 M solution of tetrabutylammonium fluoride in THF (0.67 mL, 670 μmol) was added and the solution was stirred at room temperature for 3 h. The solution was concentration *in vacuo* and the crude was purified by flash chromatography (2 to 8% MeOH in CH_2Cl_2) and then precipitation from ether and trituration with CH_2Cl_2 to remove excess tetrabutylammonium to afford nucleoside **4.8t** as a white powder (70 mg, 68%). R_f = 0.08, 8% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.61 (d, *J* = 1.2 Hz, 1H, H-C6), 5.55 (d, *J* = 3.1 Hz, 1H, H-C1′), 4.04 (dd, *J* = 5.0, 9.7 Hz, 1H, H-C4′), 3.93 (m, *J* = 2.5 Hz, 1-H, H-C4′), 3.92 (m, 1H, H-C3′), 3.23 (dd, *J* = 2.7 Hz, 1H, H-C2′), 1.76 (d, *J* = 1.2 Hz, 3H, H_3C-C5); ^{13}C NMR

(100 MHz, DMSO-d_6) δ 163.9, 150.7, 137.1, 108.3, 91.8, 75.9, 74.5, 64.0, 12.3; HRMS (m/z): $[M+H^+]^+$ calc. for $C_9H_{13}N_3O_4$, 228.0979, obs. 228.0990.

4,5-Dimethoxy-2-nitrobenzyl ((3*R*,4*R*)-2-(2-(((*E*)-(dimethylamino)methylene)amino)-6-oxo-1,6-dihydro-9*H*-purin-9-yl)-4-hydroxytetrahydrofuran-3-yl)carbamate (4.9). Nucleoside **4.8g** was dissolved in anhydrous DMF (10 mL) under argon. To this solution was added *N*,*N*-diisopropylethylamine (0.11 mL, 632 μmol) and then 4,5-dimethoxy-2-nitrobenzyl chloroformate (58 mg, 210 μmol) and the reaction was stirred at room temperature, covered from light, for 3.5 h. The reaction was quenched with MeOH (2 mL) for 1 h and then concentrated *in vacuo* and coevaporated with toluene (2x10 mL). The crude was dissolved in anhydrous DMF (5 mL) under argon and to this solution was added *N*,*N*-dimethylformamide dimethyl acetal (1 mL, 7.53 mmol) and the reaction was stirred, covered from light, at room temperature for 6 h. The reaction was concentrated *in vacuo* and coevaported with toluene (2x10 mL) and the crude was purified by flash chromatography (5 to 40% MeOH in CH_2Cl_2) to afford protected nucleoside **4.9** as a pale yellow solid (49 mg, 50%). R_f = 0.4, 10% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.35 (s, 1H, HN-N1), 8.62 (s, 1H, H-C=N), 8.08 (d, *J* = 8.2 Hz, 1H, H-N-Nvoc), 7.98 (s, 1H, H-C8), 7.70 (s, 1H, H-Ar), 7.15 (s, 1H, H-Ar), 5.82 (d, *J* = 3.9 Hz, 1H, H-C1′), 5.73 (d, *J* = 3.8 Hz, 1H, H-OC3′), 5.35 (d, *J* = 15.0 Hz, 1H, H-HC-Ar), 5.31 (d, *J* = 14.4 Hz, 1H, H-HC-Ar), 4.59 (m, 1H, H-C2′), 4.31 (m, 1H, H-C3′), 4.16-3.96 (m, 2H, H-C4′), 3.87 (s, 6H, H_3-COAr), 3.10 (s, 3H, H_3-C-N), 3.02 (s, 3H, H_3-C-N); ESI-MS (m/z): $[M+Na^+]^+$ calc. for $C_{22}H_{26}N_8O_9$, 569.2, obs. 569.1.

4,5-Dimethoxy-2-nitrobenzyl ((3*R*,4*R*)-2-(2-amino-6-oxo-1,6-dihydro-9*H*-purin-9-yl)-4-(phosphonooxy)tetrahydrofuran-3-yl)carbamate (4.10). Protected

nucleoside **4.9** was dissolved in anhydrous acetonitrile (8 mL) with 3 Å molecular sieves (10) under argon. To this solution was added 5-(ethylthio)tetrazole (20 mg, 154 μmol) and then bis(2-cyanoethyl)-*N*,*N*-diisopropyl phosphoramidite (28 mg, 103 μmol) and the reaction was stirred at room temperature for 2.7 h covered from light. To the reaction was added a 70% aqueous solution of *tert*-butyl hydroperoxide (200 μLμ, 1.45 mmol) and the reaction was stirred for another 1.3 h. Finally, the reaction was diluted with ammonium hydroxide (15 mL) and stirred at 55°C in a pressure vessel for 18 h. The reaction was concentration *in vacuo* and purified by preparative HPLC with 2 to 15% acetonitrile in 10 mM triethylammonium bicarbonate pH 7.5 over 15 min to afford 90 mg of the protected nucleotide **4.10** as a triethylammonium salt after lyophilization. Due to excess triethylammonium the actual yield of the reaction was not determined and the product was used directly in the subsequent activation reaction. ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.63 (s, 1H, H-N1), 8.35 (d, *J* = 6.9 Hz, 1H, H-N-Nvoc), 7.99 (s, 1H, H-C8), 7.69 (s, 1H, H-Ar), 7.15 (s, 1H, H-Ar), 5.73 (d, *J* = 4.8 Hz, 1H, H-C1′), 5.33 (d, *J* = 14.7 Hz, 1H, H-HC-Ar), 5.28 (d, *J* = 14.3 Hz, 1H, H-HC-Ar), 4.69 (m, 1H, H-C2′), 4.54 (m, 1H, H-C3′), 4.12 (dd, *J* = 9.1, 4.4 Hz, H-C4′), 4.07 (dd, *J* = 9.1, 5.2 Hz, H-C4′), 3.88 (s, 3H, H_3-COAr), 3.86 (s, 3H, H_3-COAr); ^{13}P NMR (160.8 MHz, DMSO-d_6) δ 0 (phosphoric acid), -1.12; ESI-MS (m/z): $[M\text{-}H^+]^-$ calc. for $C_{19}H_{22}N_7O_{12}P$, 570.1, obs. 570.1.

(3*R*,4*R*)-4-Amino-5-(2-amino-6-oxo-1,6-dihydro-9*H*-purin-9-yl)-tetrahydrofuran-3-yl hydrogen (2-methyl-1*H*-imidazol-1-yl)phosphonate (4.11). Protected nucleotide **4.10** was dissolved in a mixture of DMSO (3 mL) and DMF (2 mL) under argon and protected from light. To this solution was added 2-methylimidazole (181 mg, 2.21 mmol), triethylamine (0.1 mL, 717 μmol), triphenylphosphine (129 mg, 492 μmol), and

finally 2,2′-dipyridyl disulfide (150 mg, 681 μmol) and the reaction was stirred at room temperature for 3 h. The reaction was precipitated in a mixture of acetone (300 mL), diethyl ether (190 mL), triethylamine (20 mL) and an acetone solution saturated with sodium perchlorate (1.5 mL). The precipitate was collected by centrifugation and washed with a 1:1 acetone-ether mixture (3x40 mL). The crude pellet was dissolved in water brought to pH 10.5 with 10 N NaOH. This solution was stirred under constant 350 nm irradiation for 16.5 h. The product was purified directly from the reaction solution by preparative HPLC with 2 to 9% acetonitrile in 10 mM triethylammonium bicarbonate pH 7.5 over 12 min to afford the product as a triethylammonium salt after lyophilization. The yield of activated nucleotide **4.11** was measured by the UV spectrophotometry, assuming an extinction coefficient of 14.09 mM^{-1} cm^{-1}, and found to be 4 μmol (6% from **4.9**). ^{1}H NMR (400 MHz, D_2O) δ 7.90 (s, 1H, H-C8), 7.06 (app s, 1H, H-Im), 6.82 (app s, 1H, H-Im), 5.77 (app s, 1H, H-C1′), 4.59 (under DHO, H-C3′), 4.47 (d, *J* = 11.1 Hz, 1H, H-C4′), 4.36 (d, *J* = 10.9 Hz, 1H, H-C4′), 3.70 (app s, 1H, H-C2′), 3.21 (q, *J* = 7.3 Hz, 0.6x6H, $NH(CH_2CH_3)_3^+$), 2.20 (s, 3H, H_3-C), 3.21 (t, *J* = 7.3 Hz, 0.6x9H, $NH(CH_2CH_3)_3^+$); ^{13}C NMR (100 MHz, DMSO-d_6) δ 160.1, 158.8, 153.6, 150.3, 137.0, 121.2, 90.9, 80.2, 74.6, 61.4, 13.7; ^{31}P NMR (160.8 MHz, 10% D_2O) δ 1.15 (hydrolyzed), 0 (orthophosphate), -12.02 (activated); HRMS (m/z): $[M\text{-}H^+]^-$ calc. for $C_{13}H_{17}N_8O_5P$, 395.0987, obs. 395.0972.

Acknowledgements

The authors would like to thank Dr. John Sutherland (University of Cambridge), Dr. Brian Seed (Massachusetts General Hospital) and members of the Szostak lab for helpful discussion. The 3′-amino terminated, 5′-TAMRA labeled DNA primer was generously

provided by Dr. Sergei Gryaznov (Geron Corp.) and the TNA template was generously provided by Dr. John Chaput (Arizona State University). We thank Dr. Shao-Liang Zheng of the Harvard X-ray Crystallography Center for X-ray diffraction and analysis.

References

Altona, C, and M Sundaralingam. 1972. "Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. a New Description Using the Concept of Pseudorotation." *Journal of the American Chemical Society* 94 (August 31): 8205–8212.

Chaput, John C, and J W Szostak. 2003. "TNA Synthesis by DNA Polymerases." *Journal of the American Chemical Society* 125 (31) (August 6): 9274–9275. doi:10.1021/ja035917n.

Chaput, John C, Justin K Ichida, and J W Szostak. 2003. "DNA Polymerase-Mediated DNA Synthesis on a TNA Template." *Journal of the American Chemical Society* 125 (4) (January 29): 856–857. doi:10.1021/ja028589k.

Chen, Jesse J, Xin Cai, and J W Szostak. 2009. "N2'-->P3' Phosphoramidate Glycerol Nucleic Acid as a Potential Alternative Genetic System.." *Journal of the American Chemical Society* 131 (6) (February 18): 2119–2121. doi:10.1021/ja809069b.

Cicchi, S, M Bonanni, F Cardona, J Revuelta, and A Goti. 2003. "Indium-Mediated Reduction of Hydroxylamines to Amines." *Organic Letters* 5 (10): 1773–1776. doi:10.1021/ol0344341.

Ebert, Marc-Olivier, Christian Mang, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Bernhard Jaun. 2008. "The Structure of a TNA-TNA Complex in Solution: NMR Study of the Octamer Duplex Derived From Alpha-(L)-Threofuranosyl-(3'-2')-CGAATTCG.." *Journal of the American Chemical Society* 130 (45) (November 12): 15105–15115. doi:10.1021/ja8041959.

Eschenmoser, A. 1999. "Chemical Etiology of Nucleic Acid Structure." *Science* 284 (5423) (June 25): 2118–2124.

Ferencic, Mathias, Goverdhan Reddy, Xiaolin Wu, Sreenivasulu Guntha, Jyoti Nandy, Ramanarayanan Krishnamurthy, and Albert Eschenmoser. 2004. "Base-Pairing Systems Related to TNA Containing Phosphoramidate Linkages: Synthesis of Building Blocks and Pairing Properties.." *Chemistry & Biodiversity* 1 (7) (July): 939–979. doi:10.1002/cbdv.200490083.

Fitzsimmons, B J, Y Leblanc, and J Rokach. 1987. "[4+2] Cycloaddition of Azodicarboxylates and Glycals: a Novel and Simple Method for the Preparation of 2-Amino-2-Deoxy Carbohydrates." *Journal of the American Chemical Society* 109 (1): 285–286.

Gottlieb, H, V Kotlyar, and A Nudelman. 1997. “NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities.” *The Journal of Organic Chemistry* 62 (21) (October 17): 7512–7515.

Gustorf, von, E K, D V White, B Kim, D Hess, and J Leitich. 1970. “Photochemical and Thermal 1,2- and 1,4-Cycloaddition Reactions of Azodicarbonyl Compounds with Monoolefins.” *The Journal of Organic Chemistry* 35 (4): 1155–1165.

Heuberger, Benjamin D, and Christopher Switzer. 2006. “Nonenzymatic Oligomerization of RNA by TNA Templates.” *Organic Letters* 8 (25) (December): 5809–5811. doi:10.1021/ol062368s.

Ichida, JK, A Horhota, KY Zou, LW McLaughlin, and J W Szostak. 2005. “High Fidelity TNA Synthesis by Therminator Polymerase.” *Nucleic Acids Research* 33 (16): 5219–5225. doi:10.1093/nar/gki840.

Ichida, JK, K Zou, A Horhota, B Yu, LW McLaughlin, and J W Szostak. 2005. “An in Vitro Selection System for TNA.” *Journal of the American Chemical Society* 127 (9): 2802–2803. doi:10.1021/ja045364w.

Inoue, T, and L E Orgel. 1983. “A Nonenzymatic RNA Polymerase Model.” *Science* 219 (4586) (February 18): 859–862.

Islam, Saidul, Juan A Aguilar, Matthew W Powner, Mathias Nilsson, Gareth A Morris, and John D Sutherland. 2013. “Detection of Potential TNA and RNA Nucleoside Precursors in a Prebiotic Mixture by Pure Shift Diffusion-Ordered NMR Spectroscopy..” *Chemistry (Weinheim an Der Bergstrasse, Germany)* 19 (14) (April 2): 4586–4595. doi:10.1002/chem.201202649.

Joyce, G F, T Inoue, and L E Orgel. 1984. “Non-Enzymatic Template-Directed Synthesis on RNA Random Copolymers. Poly(C, U) Templates.” *Journal of Molecular Biology* 176 (2) (June 25): 279–306.

Kaiser, Andreas, and Clemens Richert. 2013. “Nucleotide-Based Copying of Nucleic Acid Sequences Without Enzymes.” *The Journal of Organic Chemistry* 78 (3) (February 1): 793–799. doi:10.1021/jo3025779.

Kaiser, Andreas, Sebastian Spies, Tanja Lommel, and Clemens Richert. 2012. “Template-Directed Synthesis in 3′- and 5′-Direction with Reversible Termination.” *Angewandte Chemie-International Edition in English* 51 (33) (August 13): 8299–8303. doi:10.1002/anie.201203859.

Kanavarioti, A, CF Bernasconi, D L Doodokyan, and D J Alberas. 1989. “Magnesium-Ion Catalyzed P-N Bond Hydrolysis in Imidazolide-Activated Nucleotides - Relevance to Template-Directed Synthesis of Polynucleotides.” *Journal of the American Chemical Society* 111 (18): 7247–7257.

Kim, Hyo-Joong, Alonso Ricardo, Heshan I Illangkoon, Myong Jung Kim, Matthew A Carrigan, Fabianne Frye, and Steven A Benner. 2011. "Synthesis of Carbohydrates in Mineral-Guided Prebiotic Cycles." *Journal of the American Chemical Society* 133 (24) (June 22): 9457–9468. doi:10.1021/ja201769f.

Kozlov, IA, M Zielinski, B Allart, L Kerremans, A Van Aerschot, R Busson, P Herdewijn, and L E Orgel. 2000. "Nonenzymatic Template-Directed Reactions on Altritol Oligomers, Preorganized Analogues of Oligonucleotides." *Chemistry-a European Journal* 6 (1): 151–155.

Kozlov, IA, PK Politis, A Van Aerschot, R Busson, P Herdewijn, and L E Orgel. 1999. "Nonenzymatic Synthesis of RNA and DNA Oligomers on Hexitol Nucleic Acid Templates: the Importance of the a Structure." *Journal of the American Chemical Society* 121 (12): 2653–2656.

Leblanc, Y, and B J Fitzsimmons. 1989. "[4+2] Cycloaddition Reaction of Bis (Trichloroethyl) Azodicarboxylate and Glycals: Preparation of a C1-C1 2-Amino Disaccharide." *Tetrahedron Letters* 30 (22): 2889–2892.

Leblanc, Y, B J Fitzsimmons, Springer J P, and J Rokach. 1989. "[4+2] Cycloaddition Reaction of Dibenzyl Azodicarboxylate and Glycals." *Journal of the American Chemical Society* 111 (8): 2995–3000.

Lichtenthaler, FW, T Weimer, and S Immel. 2004. "[4+2] and [2+2] Photocycloadditions of 1,2-Diketones to Glyeal and Hydroxyglycal Esters." *Tetrahedron-Asymmetry* 15 (17): 2703–2709. doi:10.1016/j.tetasy.2004.07.019.

Lincoln, Tracey A, and Gerald F Joyce. 2009. "Self-Sustained Replication of an RNA Enzyme." *Science* 323 (5918) (February 27): 1229–1232. doi:10.1126/science.1167856.

Lohrmann, R, and L E Orgel. 1976. "Template-Directed Synthesis of High Molecular Weight Polynucleotide Analogues." *Nature* 261 (5558) (May 27): 342–344.

Mansy, Sheref S, Jason P Schrum, Mathangi Krishnamurthy, Sylvia Tobé, Douglas A Treco, and J W Szostak. 2008. "Template-Directed Synthesis of a Genetic Polymer in a Model Protocell." *Nature* 454 (7200) (July 3): 122–125. doi:10.1038/nature07018.

Mineno, T, SR Choi, and MA Avery. 2002. "A Mild Deprotection of Trichloroethyl Carbamates Using Indium Metal." *Synlett* (6): 883–886.

Orgel, L E. 2000. "Origin of Life - a Simpler Nucleic Acid." *Science* 290 (5495): 1306–1307.

Pallan, Pradeep S, Christopher J Wilds, Zdzislaw Wawrzak, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Martin Egli. 2003. "Why Does TNA Cross-Pair More Strongly with RNA Than with DNA? an Answer From X-Ray Analysis.." *Angewandte Chemie-International Edition in English* 42 (47): 5893–5895. doi:10.1002/anie.200352553.

Pinheiro, Vitor B, Alexander I Taylor, Christopher Cozens, Mikhail Abramov, Marleen Renders, Su Zhang, John C Chaput, et al. 2012. “Synthetic Genetic Polymers Capable of Heredity and Evolution.” *Science* 336 (6079) (April 20): 341–344. doi:10.1126/science.1217622.

Robins, M J, R M Zou, Z Q Guo, and SF Wnuk. 1996. “Nucleic Acid Related Compounds .93. a Solution for the Historic Problem of Regioselective Sugar-Base Coupling to Produce 9-Glycosylguanines or 7-Glycosylguanines.” *The Journal of Organic Chemistry* 61 (26): 9207–9212.

Röthlingshöfer, Manuel, Eric Kervio, Tanja Lommel, Ulrich Plutowski, Annette Hochgesand, and Clemens Richert. 2008. “Chemical Primer Extension in Seconds.” *Angewandte Chemie-International Edition in English* 47 (32) (July 28): 6065–6068. doi:10.1002/anie.v47:32.

Schoning, KU, P Scholz, S Guntha, X Wu, R Krishnamurthy, and A Eschenmoser. 2000. “Chemical Etiology of Nucleic Acid Structure: the Alpha-Threofuranosyl-(3′->2′) Oligonucleotide System.” *Science* 290 (5495): 1347–1351.

Schoning, KU, P Scholz, XL Wu, S Guntha, G Delgado, R Krishnamurthy, and A Eschenmoser. 2002. “The Alpha-L-Threofuranosyl-(3′-2′)-Oligonucleotide System (‘TNA’): Synthesis and Pairing Properties.” *Helvetica Chimica Acta* 85 (12): 4111–4153.

Schrum, Jason P, Alonso Ricardo, Mathangi Krishnamurthy, J C Blain, and J W Szostak. 2009. “Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2′-Amino-2′,3′-Dideoxyribonucleoside-5′-Phosphorimidazolide Monomers.” *Journal of the American Chemical Society* 131 (40) (October 14): 14560–14570. doi:10.1021/ja906557v.

Szostak, J W, D P Bartel, and P L Luisi. 2001. “Synthesizing Life.” *Nature* 409 (6818): 387–390.

Vaidya, Nilesh, Michael L Manapat, Irene A Chen, Ramon Xulvi-Brunet, Eric J Hayden, and Niles Lehman. 2012. “Spontaneous Network Formation Among Cooperative RNA Replicators.” *Nature* 491 (7422) (November 1): 72–77. doi:10.1038/nature11549.

Valluri, M, T Mineno, RM Hindupur, and MA Avery. 2001. “Indium-Mediated Chemoselective Deprotection of Trichloroethoxycarbonyl and Trichloroacetyl Groups.” *Tetrahedron Letters* 42 (41): 7153–7154.

Wilds, CJ, Z Wawrzak, R Krishnamurthy, A Eschenmoser, and M Egli. 2002. “Crystal Structure of a B-Form DNA Duplex Containing (L)-Alpha-Threofuranosyl (3 ‘-> 2 ’) Nucleosides: a Four-Carbon Sugar Is Easily Accommodated Into the Backbone of DNA.” *Journal of the American Chemical Society* 124 (46): 13716–13721. doi:10.1021/ja0207807.

Wochner, Aniela, James Attwater, Alan Coulson, and Philipp Holliger. 2011. “Ribozyme-Catalyzed Transcription of an Active Ribozyme.” *Science* 332 (6026) (April 8): 209–212.

doi:10.1126/science.1200752.

Wright, G E, and L W Dudycz. 1984. “Synthesis and Characterization of N2-(P-N-Butylphenyl)-2‘-Deoxyguanosine and Its 5’-Triphosphate and Their Inhibition of HeLa DNA Polymerase Alpha.” *Journal of Medicinal Chemistry* 27 (2): 175–181.

Wu, T F, and L E Orgel. 1992. “Nonenzymatic Template-Directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides.” *Journal of the American Chemical Society* 114 (1): 317–322.

Wu, XL, G Delgado, R Krishnamurthy, and A Eschenmoser. 2002. “2,6-Diaminopurine in TNA: Effect on Duplex Stabilities and on the Efficiency of Template-Controlled Ligations.” *Organic Letters* 4 (8): 1283–1286. doi:10.1021/ol020016p.

Wu, XL, S Guntha, M Ferencic, R Krishnamurthy, and A Eschenmoser. 2002. “Base-Pairing Systems Related to TNA: Alpha-Threofuranosyl Oligonucleotides Containing Phosphoramidate Linkages.” *Organic Letters* 4 (8): 1279–1282. doi:10.1021/ol1020015x.

Yu, Hanyang, Su Zhang, and John C Chaput. 2012. “Darwinian Evolution of an Alternative Genetic System Provides Support for TNA as an RNA Progenitor.” *Nature Chemistry* 4 (3) (March): 183–187. doi:10.1038/nchem.1241.

Yu, Hanyang, Su Zhang, Matthew R Dunn, and John C Chaput. 2013. “An Efficient and Faithful in Vitro Replication System for Threose Nucleic Acid.” *Journal of the American Chemical Society* 135 (9) (March 6): 3583–3591. doi:10.1021/ja3118703.

Zhang, S, N Zhang, J C Blain, and J W Szostak. 2013. “Synthesis of N3′-P5′-Linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension.” *Journal of the American Chemical Society* 135 (January 7): 924–932. doi:10.1021/ja311164j.

Zhu, Ting F, and J W Szostak. 2009. “Coupled Growth and Division of Model Protocell Membranes.” *Journal of the American Chemical Society* 131 (15) (April 22): 5705–5713. doi:10.1021/ja900919c.

Chapter Five: Site-specific 'click' chemistry with RNA transcripts

J. Craig Blain, Isabel M. Vogt and Jack W. Szostak

Abstract

The site-specific modification of RNA transcripts is critical for many applications in RNA biochemistry, including fluorescent labeling, affinity tagging, chemical ligation and *in vitro* selection. Therefore, it is important to have efficient and general methods for performing these modifications. To this end, we have developed two new initiator nucleotides – derivatives of GMP with a modified 5′-phosphate that can prime transcription to make 5′-modified RNA – that contain an alkyne group for downstream 'click' chemistry to azide containing compounds. One contains a $(CH_2)_2$ linker to a terminal alkyne for copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) coupling and the other contains a bulkier dibenzylcyclooctyne group for copper-free strain-promoted azide-alkyne cycloaddition coupling (SPAAC). The CuAAC nucleotide was incorporated into 81% of transcripts when at a 10:1 ratio to GTP and the SPAAC nucleotide was incorporated into 92% at a ratio of 8:1, each without a loss in transcription yield. Both modified transcripts could be cleanly coupled to a model azide. Finally, we propose a scheme for an *in vitro* selection experiment to isolate an RNA-dependent RNA polymerase ribozyme that would make use of these initiator nucleotides.

Introduction

Chemically modified RNA oligonucleotides have many applications in biological research and biotechnology, including fluorescent labeling for *in vitro* and *in vivo* imaging (Phelps, Morris, and Beal 2012), biotin tagging for affinity chromatography (Ahsen and Noller 1995), chemical ligation (El-Sagheer and Brown 2012), and substrate conjugation for *in vitro* selection (Seelig and Jaschke 1999). Modifications can be introduced into RNA through the use of modified phosphoramidites during solid phase synthesis (Paredes, Evans, and Das 2011); however, the length of RNA that can be synthesized routinely is limited to <100 nucleotides. In contrast, *in vitro* transcription by T7 RNA polymerase is relatively inexpensive and can be used to produce RNA molecules kilobases long in milligram quantities (Studier et al. 1990; Milligan and Uhlenbeck 1989), but the transcription process is less amenable to modification. Modified nucleotide triphosphates can be incorporated throughout a transcript by RNA polymerase enzymes (Jao and Salic 2008; Langer, Waldrop, and Ward 1981; Strobel 1999) and residues can be chemically modified post-transcriptionally (Babak et al. 2004), but the count and position of modification cannot be controlled. Although non-natural, orthogonal base pairs have been used to incorporate modified nucleotide triphosphates at particular positions (Tor and Dervan 1993), this method still requires chemical synthesis of the non-natural template.

To achieve the single, site-specific modification of an RNA transcript one of three methods is typically used: (i) enzymatic ligation to a modified nucleotide or oligonucleotide using T4 RNA ligase (Kinoshita, Nishigaki, and Husimi 1997) or poly(A) polymerase (Martin and Keller 1998); (ii) oxidation of the 3′-diol to a dialdehyde and subsequent coupling to a hydrazide by reductive amination (Hansske and Cramer 1979); or (iii)

addition of an initiator nucleotide or oligonucleotide to the transcription reaction for incorporation at the 5′-terminus (Pitulle et al. 1992; Axelrod and Kramer 1985; Sampson and Uhlenbeck 1988; Huang et al. 2008). Of these methods the initiator nucleotide method is the simplest since it does not require any addition steps. An initiator nucleotide is a derivative of guanosine or adenosine 5′-monophosphate that T7 RNA polymerase can use to prime transcription, but cannot incorporate into a growing transcript since it is not activated. Although a variety of initiator nucleotides can be used (Huang 2003; Fiammengo, Musilek, and Jaschke 2005; Yisraeli and Melton 1989), guanosine 5′-monophosphorothioate provides a nucleophilic sulfur that can subsequently be coupled to a variety of electrophiles (Burgin and Pace 1990; Zhang and Cech 1997).

We sought to develop general purpose initiator nucleotides that are orthogonal to nucleophilic substitution chemistry and instead use efficient azide-alkyne 1,3-dipolar cycloaddition, or 'click', chemistry (Figure 5.1) (Kolb, Finn, and Sharpless 2001). The copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction has found broad applications in biological research due to its quickness and selectivity. However, since the reactive oxygen species that are produced by the copper(I) catalyst are damaging to RNA, conditions for efficient CuAAC coupling with RNA have been developed only recently (Hong et al. 2009; El-Sagheer and Brown 2010; Paredes and Das 2011).

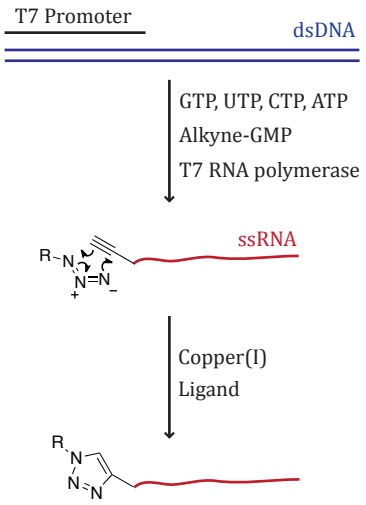


Figure 5.1. General scheme for site-specific 'click' coupling to the 5′-terminus of an RNA transcript through the initiator nucleotide method. The 5′-phosphate of GMP is modified with an alkyne moiety such that T7 RNA polymerase can use it to prime transcription, but cannot incorporate it into the transcript since it is not activated. The transcript can then be 'clicked' through an azide-alkyne cycloaddition reaction to a wide variety of functionalities (R), provided they contain an azide group.

For efficient CuAAC chemistry with RNA the copper(I) catalyst must be stabilized by a chelating ligand, usually a polytriazole like tris(benzyltriazolylmethyl)amine (TBTA) or the more water soluble tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) (Chan et al. 2004), although the less expensive ligand *N*,*N*,*N*′,*N*′′,*N*′′-pentamethyldiethylenetriamine (PMDETA) with 20% acetonitrile as a cosolvent can work just as well (Paredes and Das 2011). Better results are obtained if the copper(I) is produced *in situ* by reduction of copper(II) with sodium ascorbate after chelation (Hong et al. 2009; Paredes and Das 2011). The need for copper(I) catalysis can be removed entirely if the alkyne is strained by incorporation into an octyl ring to give a reaction known as the strain-promoted alkyne-

azide cycloaddition (SPAAC) (Agard, Prescher, and Bertozzi 2004; Ning et al. 2008), but this reaction creates a bulkier linkage and has not yet been fully exploited for RNA biochemistry (van Delft et al. 2010).

The CuAAC reaction has now been used to modify the properties of siRNA towards the development of therapeutics (Hong et al. 2009; Peacock, Fostvedt, and Beal 2010; Peacock, Maydanovych, and Beal 2010; Yamada et al. 2011; Willibald et al. 2012), to ligate oligoribonucleotides (Xu, Suzuki, and Komiyama 2009; El-Sagheer and Brown 2010; El-Sagheer and Brown 2012; Paredes and Das 2011), and to label RNA (Paredes and Das 2011; Phelps, Morris, and Beal 2012; Motorin et al. 2011). The alkyne or azide functionality is typically incorporated into the sugar or nucleobase during solid phase synthesis, but can also be incorporated enzymatically (Peacock, Fostvedt, and Beal 2010; Winz et al. 2012). Independently of our work, an initiator nucleotide with a terminal alkyne has been used to successfully label an RNA transcript (Paredes and Das 2011).

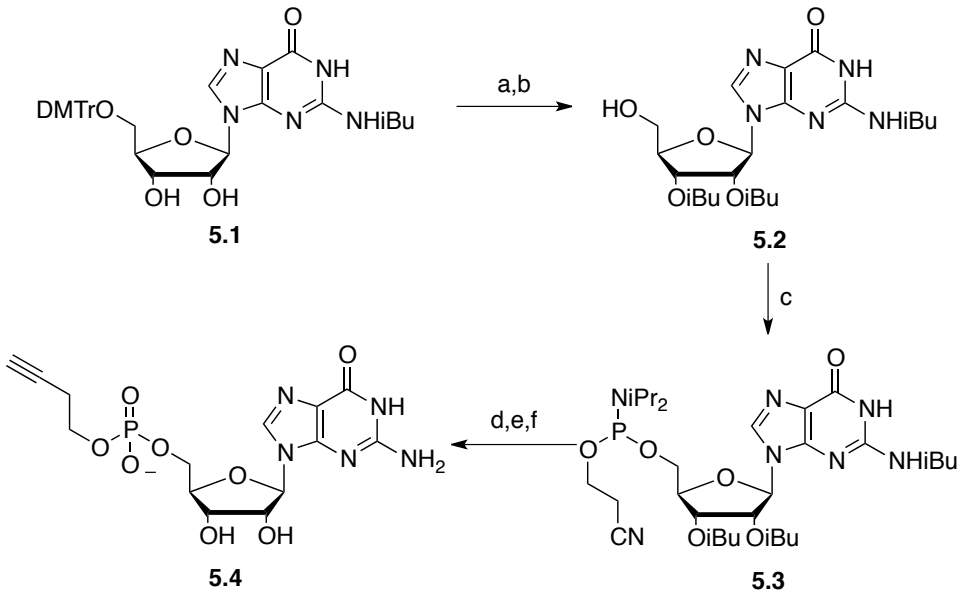
Here we report the synthesis, transcription priming and subsequent 'click' coupling of two new initiator nucleotides, one containing a terminal alkyne with a shorter linker than that of Paredes *et al.* (Paredes and Das 2011), and another containing a dibenzylcyclooctyne (DBCO) group for SPAAC coupling. These nucleotides allow for simple and efficient site-specific coupling to RNA transcripts. We will further propose a novel scheme to isolate an RNA-dependent RNA polymerase ribozyme by *in vitro* selection that makes use of these initiator nucleotides.

Results

Copper-catalyzed azide-alkyne cycloaddition on RNA transcripts

Not all derivatives of guanosine 5′-monophosphate (GMP) can be used by T7 RNA polymerase to efficiently prime transcription and there is no clear way to predict what the polymerase will accept (Fiammengo, Musilek, and Jaschke 2005). Indeed, our first attempt at using a PEG600 linker between the GMP phosphate and terminal alkyne inhibited transcription with minimal incorporation. Therefore, we next synthesized an initiator nucleotide with a much shorter $(CH_2)_2$ linker, giving GMP derivative **5.4**. We synthesized this nucleotide through simple phosphoramidite chemistry after protecting the nucleobase and 2′- and 3′-hydroxyl groups (Scheme 5.1). Although we synthesized protected nucleoside **5.2** in a simple two-step procedure, we could have also used commercially available **5.5**. Initiator nucleotide **5.4** was HPLC purified prior to use.

Scheme 5.1. Synthesis of an initiator nucleotide with a terminal alkyne.[a]



Scheme 5.1 (continued). [a]Reagents and conditions: (a) isobutyric anhydride, pyridine, room temperature, 16 h; (b) trichloroacetic acid, CH_2Cl_2, room temperature, 1 h, 63%; (c) 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, *N*,*N*-diisopropylethylamine, CH_2Cl_2, room temperature, 2.5 h, 86%; (d) 3-butyn-1-ol, 5-(ethylthio)tetrazole, MeCN, room temperature, 1.5h; (e) *tert*-butyl hydroperoxide, room temperature, 1 h; (f) ammonium hydroxide, 60°C, 16 h, 63%.

To test the incorporation and conjugation efficiency of initiator nucleotide **5.4** we used the same synthetic double-stranded DNA sequence used by Jäcshke and colleagues (Fiammengo, Musilek, and Jaschke 2005) with a standard class III φ6.5 promoter sequence and a 25 nucleotide product for easy polyacrylamide gel electrophoresis (PAGE) analysis. Based on their results, we used 0.4 mM GTP and 1 mM of each other nucleotide triphosphate to maximize the incorporation ratio. With these conditions we found that 4 mM of **5.4** increased the transcription yield 1.4±0.3 times (mean±SE, n=2), although the transcripts could not be resolved to determine the incorporation efficiency. We proceeded with the CuAAC coupling reaction, using azide-PEG_{10}-amine as a model azide for the gel shift assay. We used coupling conditions similar to those of Das and colleagues (Paredes and Das 2011), with 15 mM $CuSO_4$, 30 mM THPTA, and 15% acetonitrile, buffered to pH 7.5 with TrisHCl and the reaction was initiated by the addition of 15 mM sodium ascorbate. The concentrations of azide and RNA were 100 μM and roughly 5 μM, respectively. The reaction fit well to a single exponential with a pseudo-first order rate constant of 2.6 h^{-1} and a plateau of 19% (Figure 5.2). The plateau likely represents the fraction of transcripts that did not contain the initiator nucleotide.

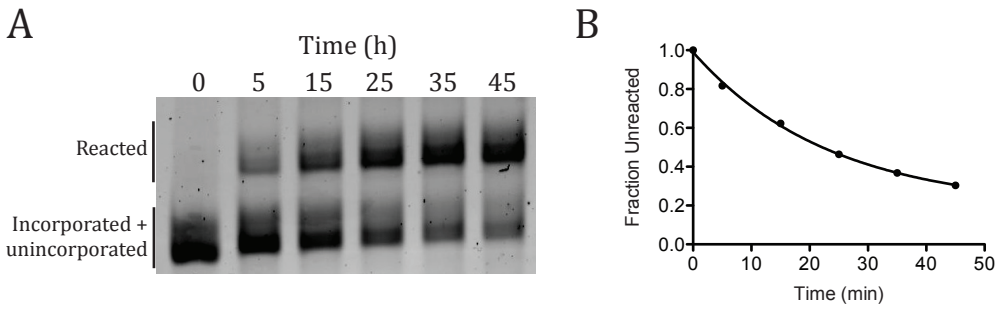


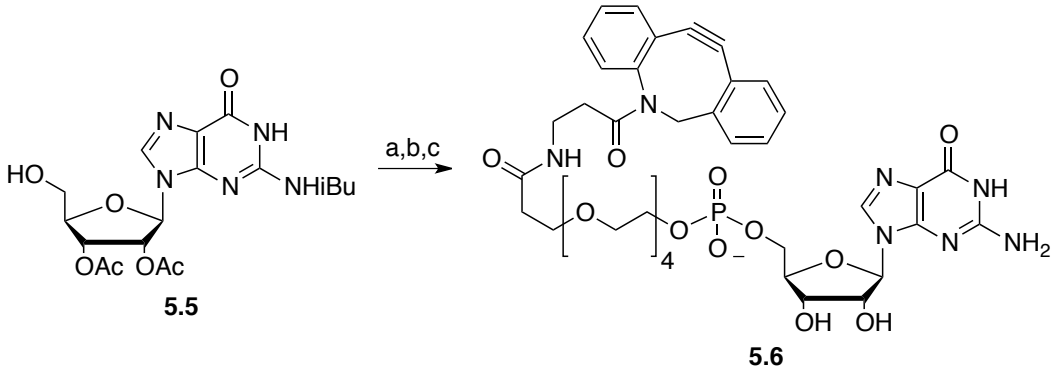
Figure 5.2. Site-specific CuAAC coupling to an RNA transcript. (A) PAGE analysis of CuAAC coupling between azide-PEG_{10}-amine and a 25 nucleotide transcript primed with initiator nucleotide **5.4**. The coupling reaction contained 15 mM $CuSO_4$, 30 mM THPTA, and 15% MeCN, buffered to pH 7.5 with TrisHCl and was initiated by the addition of 15 mM sodium ascorbate. The unreacted transcripts that were primed by initiator nucleotide **5.4** versus GTP were not resolved. (B) Quantitation of the PAGE image. The plot was fit to a single exponential decay with k_{obs} of 2.6 h^{-1} and a plateau of 19%.

Strain-promoted azide-alkyne cycloaddition on RNA transcripts

After obtaining promising results with the site-specific CuAAC coupling to RNA, we moved on to develop an initiator nucleotide that did not require copper(I) catalysis and instead relied on SPAAC coupling. Since strained alkynes in octyl rings are bulkier than terminal alkynes, efficient incorporation of such an initiator nucleotide by T7 RNA polymerase was less certain. Since a DBCO-PEG_4-phosphoramidite is commercially available, we synthesized DBCO-PEG_4-GMP (**5.6**) in a one-pot three-step procedure according to Scheme 5.2. The overall yield was low at 44%, but a single synthesis was still

able to provide 15 μmol of **5.6** and a 100 μmol transcription reaction only requires 0.4 μmol.

Scheme 5.2. Synthesis of an initiator nucleotide with a strained alkyne.[a]



[a]Reagents and conditions: (a) dibenzylcyclooctyl-PEG$_4$-phosphoramidite, 5-(ethylthio)tetrazole, MeCN, room temperature, 2.5 h; (b) *tert*-butyl hydroperoxide, room temperature, 2.25 h; (c) ammonium hydroxide, 60°C, 17 h, 44%.

Fortunately, we found that initiator nucleotide **5.6** did not inhibit transcription and was incorporated into 92% of transcripts when at a concentration of 3.2 mM using the same transcription conditions described above (Figure 5.3A). The copper-free SPAAC coupling was also effective with 97% conversion to the azide-PEG$_{10}$-amine-conjugated RNA after 3 hours (Figure 5.3B). The coupling rate was slower than the CuAAC reaction with a pseudo-first order rate constant of 0.90 h^{-1} at the same concentrations (Figure 5.3C), but could potentially be increased by using higher concentrations.

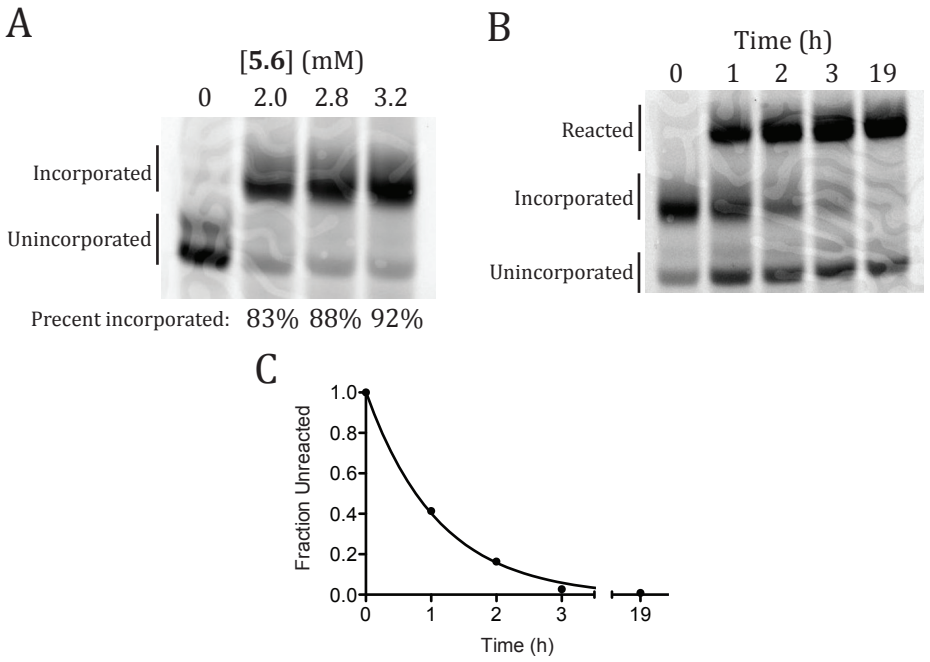


Figure 5.3. Site-specific copper-free SPAAC coupling to an RNA transcript. (A) PAGE analysis of RNA from transcription reactions containing 0.4 mM GTP and the indicated amount of initiator nucleotide **5.6**. (B) PAGE analysis of the SPAAC coupling reaction on transcripts primed with **5.6**. (C) Quantitation of the SPAAC coupling reaction. The plot was fit to a single exponential decay with a k_{obs} of 0.90 h^{-1}.

Discussion

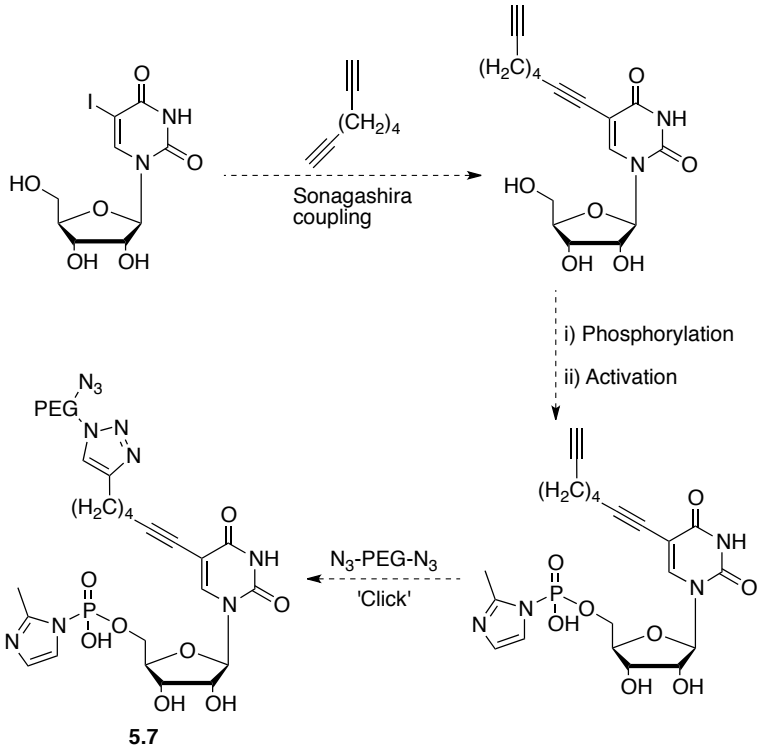
We have developed two new initiator nucleotides containing alkyne groups that T7 RNA polymerase can use to prime transcription, resulting in RNA transcripts that can be coupled to azides using 'click' chemistry. One of these nucleotides (**5.4**) contains a short linker to a terminal alkyne for CuAAC coupling, and the other (**5.6**) contains a strained DBCO alkyne for copper-free SPAAC coupling. These nucleotides can be easily synthesized (Schemes 5.1 and 5.2), incorporated into transcription reactions and coupled to azides (Figures 5.2 and 5.3). Overall, the procedures outlined here are some of the simplest and

most efficient for site-specific coupling to RNA transcripts. Azide groups can be easily installed in compounds by nucleophilic substitution and, due to the popularity of click chemistry, many azide-functionalized compounds are commercially available. Efficient and orthogonal copper-free coupling can also be achieved by tetrazine-alkene cycloaddition chemistry, and this reaction has already been applied to nucleic acids (Schoch, Wiessler, and Jäschke 2010; Seckute, Yang, and Devaraj 2013). Initiator nucleotides containing tetrazine, norbornene or cyclopentene functionality could therefore also be useful furture developments.

Although the initiator nucleotides presented here have broad potential applications, our main interest was in facilitating the *in vitro* selection of ribozymes. These experiments often involve coupling a transcribed RNA pool to a substrate (Seelig and Jaschke 1999; Coleman and Huang 2002; Zhang and Cech 1997) and it is laborious to develop a new coupling method for each selection. Specifically, we propose coupling the transcribed RNA pool to a 2-methylimidazole-activate uridine nucleotide with a PEG-linker (Scheme 5.3). The RNA pool can then be selected for template-directed nucleotide ligation activity to the 3′-terminus of a primer-template hairpin conjugated to biotin as a model for RNA-dependent RNA polymerase activity (Figure 5.4). Optimization of the R18 polymerase has been met with limited success (Wochner et al. 2011), possibly since it was developed from a catalytic core that was originally selected as an oligonucleotide ligase that has little affinity for a primer-template complex and requires a relatively unreactive triphosphate substrate (Lawrence and Bartel 2003). The selection proposed here would require substrate affinity from the ribozyme and the chemical reaction being catalyzed is more favorable (Chapter 2). A requirement for fidelity could be incorporated into the selection

by the use of a negative selection with a mismatched template. Similarly, processivity could be explicitly selected for by requiring the addition of several activated nucleotides from solution before addition of the conjugated nucleotide.

Scheme 5.3. Proposed synthesis for an azide-PEG coupled, activated uridine nucleotide.



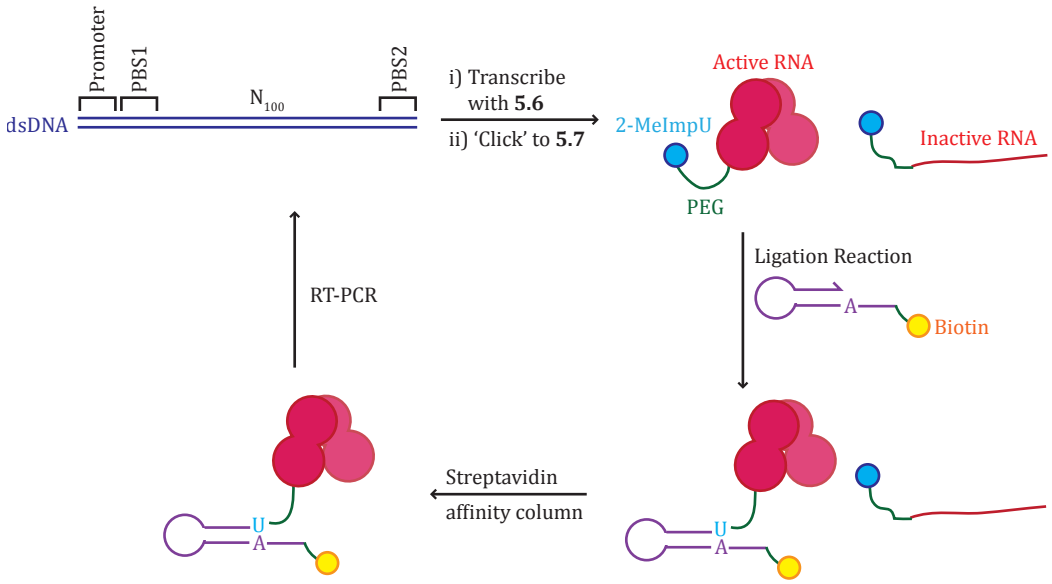


Figure 5.4. Proposed *in vitro* selection experiment to isolate a new RNA-dependent RNA polymerase ribozyme core that uses 2-methylimidazole-activate substrates.

Methods and Materials

General

All reagents were purchased from Sigma-Aldrich unless stated otherwise. Deuterated solvents were purchased from Cambridge Isotope Laboratories. DNA oligonucleotides were purchased from IDT DNA. Flash chromatography was performed on a Biotage SP1 instrument with HP-Sil columns. NMR spectroscopy was performed on a 400 MHz Varian spectrometer (Oxford AS-400) operating at 25°C. Spectra were referenced to the solvent peak according to published values (Gottlieb, Kotlyar, and Nudelman 1997), except for ^{31}P spectra, which were referenced to orthophosphate (0 ppm). All gels were imaged on a Typhoon Scanner 9400 (GE Healthcare) and analyzed using ImageQuant TL software (GE Healthcare) with the standard Cy3 settings and a PMT of 600. All further analysis was performed with Excel (Microsoft) and Prism (GraphPad). THPTA was

synthesized according to published procedures, except with a benzoyl protecting group instead of acetyl (Hong et al. 2009). Low resolution ESI-MS was performed on a Bruker Esquire 6000 with direct injection. Preparative HPLC purification was performed on a Varian ProStar instrument with a Varian Dynamax Microsorb C18 column (250x21.4 mm).

Initiator nucleotide primed transcription with T7 RNA polymerase

The conditions for the click reaction were as follows: 1 mM ATP, UTP and CTP, 0.4 mM GTP, 1 μM sense DNA strand (5′-TCT AAT ACG ACT CAC TAT AGG AGC TCA GCC TAC GAG CCT GAG CC-3′, with the transcribed sequence underlined), 1μM antisense DNA strand (5′-GGC TCA GGC TCG TAG GCT GAG CTC CTA TAG TGA GTC GTA TTA GA-3′), 5 μM Cy3-UTP (Perkin-Elmer, for visualization), the indicated concentration of initiator nucleotide **5.4** or **5.6**, and 2 U/μL T7 RNA Polymerase (New England Biolabs) in 1X T7 buffer (New England Biolabs). The total reaction volume ranged from 25 to 150 μL. The polymerase was added last and the reactions were incubated at 37°C for 16 h. Removal of excess initiator nucleotide was critical for the subsequent 'click' reaction, therefore the transcripts were purified by miRNeasy column purification (Qiagen) according to the standard protocol.

Copper(I)-catalyzed azide-alkyne cycloaddition

The conditions for the CuAAC coupling were as follows: 100 mM TrisHCl pH 7.5, 15 mM $CuSO_4$, 30 mM THPTA, 15% MeCN, 100 μM N_3-PEG_{10}-NH_2, roughly 5 μM transcript, and 15 mM sodium ascorbate at 20°C with a total concentration of 500 μL. The ascorbate was added last to reduce copper(II) to copper(I) and begin the reaction. At 5, 15, 25, 35, and 45 min a 100 μL was taken and ethanol precipitated. The samples were analyzed by urea-TBE

20% denaturing PAGE. The intensity of each peak was normalized to the total integration for all bands in that lane.

Strain-promoted azide-alkyne cycloaddition

The conditions for the SPAAC coupling were as follows: 100 mM TrisHCl pH 7.5, 100 μM N_3-PEG_{10}-NH_2, and roughly 5 μM transcript at 20°C with a total concentration of 500 μL. Samples were analyzed as per the CuAAC coupling, but with time points of 1, 2, 3, and 19 h.

Synthetic procedures

2′-*O*,3′-*O*,*N*^{2}-Triisobutyrylguanosine (5.2). Protected guanosine **5.1** (5.17 g, 7.89 mmol) was dissolved in anhydrous pyridine (70 mL) under argon. Isobutyric anhydride (5.0 mL, 29.2 mmol) was added slowly and the reaction was stirred at room temperature for 16 h. The solution was concentrated *in vacuo* and the residue was partitioned between EtOAc (300 mL) and 5% $NaHCO_3$ (200 mL) and the organic layer was then washed with water (200 mL) and brine (200 mL), dried over $MgSO_4$, filtered, and concentrated *in vacuo* to a crude foam. This crude was dissolved in a solution of 5% trichloroacetic acid in CH_2Cl_2 (50 mL) and stirred at room temperature for 1 h under argon. The reaction was neutralized by the addition of 5% $NaHCO_3$ (50 mL) and then the CH_2Cl_2 phase was washed with water (50 mL) and brine (50 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo* to a crude foam. The crude was purified by flash chromatography (0 to 6% MeOH in CH_2Cl_2) to afford protected nucleoside **5.2** as a white foam (2.46 g, 63%). R_f = 0.24, 6% MeOH in CH_2Cl_2; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.71 (s, 1H, H-C8), 5.95-5.88 (m, 2H, H-C1′ and H-C2′), 5.67 (dd, *J* = 5.1, 2.5 Hz, H-C3′), 5.01 (d, *J* = 8.3 Hz, 1H, H-OC5′), 4.27 (d, *J* = 1.9 Hz, 1H, H-C4′), 3.96 (dd, *J*

= 12.6, 2.0 Hz, 1H, H-C5′), 3.83 (dd, *J* = 1.9, 1.9 Hz, 1H, H-C5′), 2.71 (7, *J* = 6.9 Hz, 1H, H-C(CH_3)$_2$), 2.62 (7, *J* = 7.0 Hz, 1H, H-C(CH_3)$_2$), 2.51 (7, *J* = 7.0 Hz, 1H, H-C(CH_3)$_2$), 1.27 (d, *J* = 6.9 Hz, 3H, C*H*$_3$), 1.27 (d, *J* = 6.9 Hz, 3H, C*H*$_3$)), 1.21 (d, *J* = 7.0 Hz, 6H, 2xC*H*$_3$)), 1.12 (d, *J* = 7.0 Hz, 3H, C*H*$_3$)), 1.09 (d, *J* = 7.0 Hz, 3H, C*H*$_3$); ESI-MS (m/z): $[M-H^+]^-$ calc. for $C_{22}H_{31}N_5O_8$, 492.2, obs. 492.2.

2′-*O*,3′-*O*,*N*^{2}-Triisobutyrylguanosine-5′-*O*-[*O*-(2-cyanoethyl)-*N*,*N*-diisopropyl-phosphamidite] (5.3). Protected nucleoside **5.2** (516 mg, 1.05 mmol) was dissolved in anhydrous CH_2Cl_2 (11 mL) under argon. The solution was cooled to 0°C and then *N*,*N*-diisopropylethylamine (0.58 mL, 3.36 mmol) and 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (500 mg, 2.10 mmol) were added. The reaction was gradually warmed to room temperature and stirred for 2.5 h. The solution was concentrated *in vacuo* and the residue was purified by flash chromatography (0 to 5% MeOH in CH_2Cl_2) to afford phosphoramidite **5.3** as a white foam (628 mg, 86%). R_f = 0.42, 8% MeOH in CH_2Cl_2; Note: complex spectrum due to diastereomers ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.22 (s, 1H, H-C8), 8.20 (s, 1H, H-C8), 6.07 (d, *J* = 7.3 Hz, 2H, H-C1′), 5.82-5.77 (m, 2H, H-C2′), 5.52 (dd, *J* = 5.5, 2.4 Hz, 1H, H-C3′), 5.48 (dd, *J* = 5.5, 2.8 Hz, 1H, H-C3′), 4.36 (m, 2H, H-C4′), 4.1-3.4 (m, H-C5′ and CH_2CH_2), 2.90-2.45 (m, H-C(CH_3)$_2$), 1.20-0.97 (m, 60H, C*H*$_3$); ESI-MS (m/z): $[M+H^+]^+$ calc. for $C_{31}H_{48}N_7O_9P$, 694.3, obs. 694.2.

Guanosine-5′-*O*-(but-3-yn-1-yl phosphate) (5.4). Phosphoramidite **5.3** was dissolved in anhydrous MeCN (10 mL) under argon. 3-Butyn-1-ol (7.9 mL, 102 μmol) and 5-(ethylthio)tetrazole (8.0 mg, 55 μmol) were added to the solution and the reaction was stirred at room temperature for 1.5 h. To the reaction was added a 70% solution of *tert*-butyl hydroperoxide (128 μL, 924 μL) and stirring was continued at room temperature for

another 1 h. Finally, The reaction was transferred to a pressure vessel, diluted with ammonium hydroxide (10 mL) and then stirred at 60°C for 16 h. The solution was concentrated *in vacuo* and the product was purified by preparative HPLC with 2 to 20% acetonitrile in 10 mM triethylammonium bicarbonate at pH 7.5 over 20 min to afford the product as triethylammonium salt after lyophilization (63% yield by HPLC). ^{1}H NMR (400 MHz, D_2O) δ 8.04 (s, 1H, H-C8), 5.91 (d, *J* = 5.7 Hz, 1H, H-C1′), 4.82 (m, H-C2′), 4.50 (m, 1H, H-C3′), 4.31 (br s, 1H, H-C4′), 4.08 (br s, 2H, H-C5′), 3,79 (dt, *J* = 6.6, 6.6 Hz, 2H, C*H*$_2$CH$_2$CCH), 3.18 (q, *J* = 7.3 Hz, 1.2x6H, NH(C*H*$_2$CH$_3$)$_3^+$), 2.40 (t, *J* = 5.9 Hz, 2H, CH$_2$C*H*$_2$CCH), 2.26 (s, 1H, H-CC), 1.26 (t, *J* = 7.3 Hz, 1.2x9H, NH(CH$_2$C*H*$_3$)$_3^+$); ^{13}C NMR (100 MHz, D_2O) δ 159.6, 154.6, 152.4, 138.0, 117.0, 116.6, 87.5, 84.2, 73.8, 70.9, 65.4, 64.3, 47.2, 27.7, 20.3, 8.7; ^{31}P NMR (160.8 MHz, D_2O) δ 0 (orthophosphate), -2.31; ESI-MS (m/z): $[M-H^+]^-$ calc. for $C_{14}H_{18}N_5O_8P$, 414.1, obs. 414.1.

Guanosine-5′-*O*-(dibenzylcyclooctyl-PEG4 phosphate) (5.6). 2′,3′-*O*-Acetyl-*N*^{2}-isobutyrylguanosine (ChemGenes) (16 mg, 37 μmol) was dissolved in anhydrous MeCN (5 mL) under argon with 3 Å molecular sieves (10). Dibenzylcyclooctyl-PEG4-phosphoramidite (Click Chemistry Tools) (25 mg, 34 μmol) and 5-(ethylthio)tetrazole (5 mg, 38 μmol) were added to the reaction and then it was stirred at room temperature for 2.5 h. To this solution was added a 70% solution of *tert*-butyl hydroperoxide (50 μL, 350 μmol) and stirring was continued for 2.25 h. Finally, the reaction was transferred to a pressure vessel and diluted with ammonium hydroxide (20 mL) and then stirred at 60°C for 17 h. The reaction was concentrated *in vacuo* and the crude was purified by preparative HPLC with 2 to 35% acetonitrile in 10 mM triethylammonium bicarbonate pH 7.5 over 30 min to afford the product as a triethylammonium salt after lyophilization (44% yield by

HPLC). ^{1}H NMR (400 MHz, D_2O) δ 7.94 (s, 1H, H-C8), 7.64-7.27 (m, 8H, H-Ar), 5.83 (t, *J* = 5.3 Hz, 1H, H-C1′), 3.75 (d, *J* = 14.3 Hz, 1H, H-octyl), 4.66 (dt, *J* = 6.0, 6.0 Hz, 1H, H-C2′), 4.43 (dt, *J* = 4.3, 4.3 Hz, 1H, H-C3′), 4.23 (app d, *J* = 10.0 Hz, 1H, H-C4′), 4.05 (m, 2H, H-C5′), 3.90 (m, 2H, C*H*$_2$-OPO_3), 3.75 (dd, *J* = 14.3, 3.7 Hz, 1H, H-octyl), 3.60-3.49 (m, 18H, C*H*$_2$), 3.07 (q, *J* = 7.3 Hz, 3.2x6H, $NH(CH_2CH_3)_3^+$), 2.4-2.15 (m, 4H, C*H*$_2$), 1.22 (t, *J* = 7.3 Hz, 3.2x9H, $NH(CH_2CH_3)_3^+$); ^{13}C NMR (100 MHz, D_2O) δ 174.2, 152.2, 151.0, 148.2, 136.7, 132.4, 129.7, 129.5, 129.0, 128.7, 127.6, 126.3, 122.8, 122.1, 117.4, 114.8, 108.4, 87.1, 84.0, 74.3, 70.9, 70.6, 70.1, 67.2, 65.5, 65.3, 56.0, 47.2, 36.5, 36.4, 34.3, 8.9; ^{31}P NMR (160.8 MHz, D_2o) δ 0 (orthophosphate), -2.28; ESI-MS (m/z): $[M-H^+]^-$ calc. for $C_{39}H_{48}N_7O_{14}P$, 868.3, obs. 868.2.

References

Agard, Nicholas J, Jennifer A Prescher, and Carolyn R Bertozzi. 2004. “A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems.” *Journal of the American Chemical Society* 126 (46) (November 24): 15046–15047. doi:10.1021/ja044996f.

Ahsen, von, U, and H F Noller. 1995. “Identification of Bases in 16S rRNA Essential for tRNA Binding at the 30S Ribosomal P Site.” *Science* 267 (5195) (January 13): 234–237.

Axelrod, V D, and F R Kramer. 1985. “Transcription From Bacteriophage T7 and SP6 RNA Polymerase Promoters in the Presence of 3′-Deoxyribonucleoside 5′-Triphosphate Chain Terminators.” *Biochemistry* 24 (21) (October 8): 5716–5723.

Babak, T, W Zhang, Q Morris, B J Blencowe, and T R Hughes. 2004. “Probing microRNAs with Microarrays: Tissue Specificity and Functional Inference.” *Rna* 10 (11) (November 1): 1813–1819. doi:10.1261/rna.7119904.

Burgin, A B, and N R Pace. 1990. “Mapping the Active Site of Ribonuclease P RNA Using a Substrate Containing a Photoaffinity Agent.” *Embo Journal* 9 (12) (December): 4111–4118.

Chan, Timothy R, Robert Hilgraf, K Barry Sharpless, and Valery V Fokin. 2004. “Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis.” *Organic Letters* 6 (17) (August 19): 2853–2855. doi:10.1021/ol0493094.

Coleman, Tricia M, and Faqing Huang. 2002. “RNA-Catalyzed Thioester Synthesis.” *Chemistry & Biology* 9 (11) (November): 1227–1236.

El-Sagheer, Afaf H, and Tom Brown. 2010. “New Strategy for the Synthesis of Chemically Modified RNA Constructs Exemplified by Hairpin and Hammerhead Ribozymes.” *Proceedings of the National Academy of Sciences USA* 107 (35) (August 31): 15329–15334. doi:10.1073/pnas.1006447107/-/DCSupplemental.

El-Sagheer, Afaf H, and Tom Brown. 2012. “Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology.” *Accounts of Chemical Research* 45 (8) (August 21): 1258–1267. doi:10.1021/ar200321n.

Fiammengo, R, K Musilek, and A Jaschke. 2005. “Efficient Preparation of Organic Substrate-RNA Conjugates via in Vitro Transcription.” *Journal of the American Chemical Society* 127 (25): 9271–9276. doi:10.1021/ja051179m.

Gottlieb, H, V Kotlyar, and A Nudelman. 1997. “NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities.” *Journal of Organic Chemistry* 62 (21) (October 17): 7512–7515.

Hansske, F, and F Cramer. 1979. “Modification of the 3′ Terminus of tRNA by Periodate Oxidation and Subsequent Reaction with Hydrazides.” *Methods in Enzymology* 59: 172–181.

Hong, Vu, Stanislav I Presolski, Celia Ma, and M G Finn. 2009. “Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation.” *Angewandte Chemie-International Edition in English* 48 (52): 9879–9883. doi:10.1002/anie.200905087.

Huang, F. 2003. “Efficient Incorporation of CoA, NAD and FAD Into RNA by in Vitro Transcription.” *Nucleic Acids Research* 31 (3) (February 1): 8e–8. doi:10.1093/nar/gng008.

Huang, Faqing, Jun He, Yilin Zhang, and Yanlin Guo. 2008. “Synthesis of Biotin-AMP Conjugate for 5 ' Biotin Labeling of RNA Through One-Step in Vitro Transcription.” *Nature Protocols* 3 (12): 1848–1861. doi:10.1038/nprot.2008.185.

Jao, Cindy Y, and Adrian Salic. 2008. “Exploring RNA Transcription and Turnover in Vivo by Using Click Chemistry.” *Proceedings of the National Academy of Sciences USA* 105 (41) (October 14): 15779–15784. doi:10.1073/pnas.0808480105.

Kinoshita, Y, K Nishigaki, and Y Husimi. 1997. “Fluorescence-, Isotope- or Biotin-Labeling of the 5′-End of Single-Stranded DNA/RNA Using T4 RNA Ligase.” *Nucleic Acids Research* 25 (18) (September 15): 3747–3748.

Kolb, H, M Finn, and K Sharpless. 2001. “Click Chemistry: Diverse Chemical Function From a Few Good Reactions.” *Angewandte Chemie-International Edition in English* 40 (11) (June

1): 2004–2021.

Langer, P R, A A Waldrop, and D C Ward. 1981. “Enzymatic Synthesis of Biotin-Labeled Polynucleotides: Novel Nucleic Acid Affinity Probes.” *Proceedings of the National Academy of Sciences USA* 78 (11) (November): 6633–6637.

Lawrence, MS, and D P Bartel. 2003. “Processivity of Ribozyme-Catalyzed RNA Polymerization.” *Biochemistry* 42 (29): 8748–8755. doi:10.1021/bi0342281.

Martin, G, and W Keller. 1998. “Tailing and 3′-End Labeling of RNA with Yeast Poly(a) Polymerase and Various Nucleotides.” *Rna* 4 (2) (February): 226–230.

Milligan, J F, and O C Uhlenbeck. 1989. “Synthesis of Small RNAs Using T7 RNA Polymerase.” *Methods in Enzymology* 180: 51–62.

Motorin, Y, J Burhenne, R Teimer, K Koynov, S Willnow, E Weinhold, and M Helm. 2011. “Expanding the Chemical Scope of RNA:Methyltransferases to Site-Specific Alkynylation of RNA for Click Labeling.” *Nucleic Acids Research* 39 (5) (March 19): 1943–1952. doi:10.1093/nar/gkq825.

Ning, Xinghai, Jun Guo, Margreet A Wolfert, and Geert-Jan Boons. 2008. “Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions.” *Angewandte Chemie-International Edition in English* 47 (12): 2253–2255. doi:10.1002/anie.200705456.

Paredes, Eduardo, and Subha R Das. 2011. “Click Chemistry for Rapid Labeling and Ligation of RNA.” *Chembiochem* 12 (1) (January 3): 125–131. doi:10.1002/cbic.201000466.

Paredes, Eduardo, Molly Evans, and Subha R Das. 2011. “RNA Labeling, Conjugation and Ligation.” *Methods* 54 (2) (June 1): 251–259. doi:10.1016/j.ymeth.2011.02.008.

Peacock, Hayden, Erik Fostvedt, and Peter A Beal. 2010. “Minor-Groove-Modulating Adenosine Replacements Control Protein Binding and RNAi Activity in siRNAs.” *ACS Chemical Biology* 5 (12) (December 17): 1115–1124. doi:10.1021/cb100245u.

Peacock, Hayden, Olena Maydanovych, and Peter A Beal. 2010. “N2-Modified 2-Aminopurine Ribonucleosides as Minor-Groove-Modulating Adenosine Replacements in Duplex RNA.” *Organic Letters* 12 (5) (March 5): 1044–1047. doi:10.1021/ol100019r.

Phelps, Kelly, Alexi Morris, and Peter A Beal. 2012. “Novel Modifications in RNA.” *ACS Chemical Biology* 7 (1) (January 20): 100–109. doi:10.1021/cb200422t.

Pitulle, C, R G Kleineidam, B Sproat, and G Krupp. 1992. “Initiator Oligonucleotides for the Combination of Chemical and Enzymatic RNA Synthesis.” *Gene* 112 (1) (March 1): 101–105.

Sampson, J R, and O C Uhlenbeck. 1988. “Biochemical and Physical Characterization of an

Unmodified Yeast Phenylalanine Transfer RNA Transcribed in Vitro." *Proceedings of the National Academy of Sciences USA* 85 (4): 1033–1037.

Schoch, Juliane, Manfred Wiessler, and Andres Jäschke. 2010. "Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels-Alder Reaction." *Journal of the American Chemical Society* 132 (26) (July 7): 8846–8847. doi:10.1021/ja102871p.

Seckute, Jolita, Jun Yang, and Neal K Devaraj. 2013. "Rapid Oligonucleotide-Templated Fluorogenic Tetrazine Ligations." *Nucleic Acids Research* (June 17). doi:10.1093/nar/gkt540.

Seelig, B, and A Jaschke. 1999. "A Small Catalytic RNA Motif with Diels-Alderase Activity." *Chemistry & Biology* 6 (3): 167–176.

Strobel, S A. 1999. "A Chemogenetic Approach to RNA Function/Structure Analysis." *Current Opinion in Structural Biology* 9 (3) (June): 346–352. doi:10.1016/S0959-440X(99)80046-3.

Studier, F W, A H Rosenberg, J J Dunn, and J W Dubendorff. 1990. "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes." *Methods in Enzymology* 185: 60–89.

Tor, Y, and P B Dervan. 1993. "Site-Specific Enzymatic Incorporation of an Unnatural Base, N6-(6-Aminohexyl)Isoguanine, Into RNA." *Journal of the American Chemical Society* 115 (May 31): 4461–4467.

van Delft, Pieter, Nico J Meeuwenoord, Sascha Hoogendoorn, Jasper Dinkelaar, Herman S Overkleeft, Gijsbert A van der Marel, and Dmitri V Filippov. 2010. "Synthesis of Oligoribonucleic Acid Conjugates Using a Cyclooctyne Phosphoramidite." *Organic Letters* 12 (23) (December 3): 5486–5489. doi:10.1021/ol102357u.

Willibald, Julian, Johannes Harder, Konstantin Sparrer, Karl-Klaus Conzelmann, and Thomas Carell. 2012. "Click-Modified Anandamide siRNA Enables Delivery and Gene Silencing in Neuronal and Immune Cells." *Journal of the American Chemical Society* (July 23): 120723103008002. doi:10.1021/ja303251f.

Winz, M L, A Samanta, D Benzinger, and A Jaschke. 2012. "Site-Specific Terminal and Internal Labeling of RNA by Poly(a) Polymerase Tailing and Copper-Catalyzed or Copper-Free Strain-Promoted Click Chemistry." *Nucleic Acids Research* 40 (10) (May 21): e78–e78. doi:10.1093/nar/gks062.

Wochner, Aniela, James Attwater, Alan Coulson, and Philipp Holliger. 2011. "Ribozyme-Catalyzed Transcription of an Active Ribozyme." *Science* 332 (6026) (April 8): 209–212. doi:10.1126/science.1200752.

Xu, Yan, Yuta Suzuki, and Makoto Komiyama. 2009. "Click Chemistry for the Identification of G-Quadruplex Structures: Discovery of a DNA-RNA G-Quadruplex." *Angewandte Chemie-International Edition in English* 48 (18): 3281–3284.

doi:10.1002/anie.200806306.

Yamada, Takeshi, Chang Geng Peng, Shigeo Matsuda, Haripriya Addepalli, K Narayanannair Jayaprakash, Md Rowshon Alam, Kathy Mills, et al. 2011. “Versatile Site-Specific Conjugation of Small Molecules to siRNA Using Click Chemistry.” *Journal of Organic Chemistry* 76 (5) (March 4): 1198–1211. doi:10.1021/jo101761g.

Yisraeli, J K, and D A Melton. 1989. “Synthesis of Long, Capped Transcripts in Vitro by SP6 and T7 RNA Polymerases.” *Methods in Enzymology* 180: 42–50.

Zhang, B, and T R Cech. 1997. “Peptide Bond Formation by in Vitro Selected Ribozymes.” *Nature* 390 (6655) (November 6): 96–100. doi:10.1038/36375.

Chapter Six: Concluding remarks

J. Craig Blain

Summary of Findings

Throughout this dissertation we have discussed different approaches to the development of self-replicating nucleic acids as the genetic material for a protocell. We began in chapter one with an overview of the current state of protocell research, covering non-enzymatic and ribozyme-catalyzed nucleic acid self-replication, minimal systems for protein enzyme catalyzed genome replication, and self-replicating compartments. Chapter two reported recent results on the chemical mechanism of non-enzymatic primer extension reactions with 2-methylimidazole-activated ribonucleotides. This reaction has become the standard model for non-enzymatic RNA replication since it was developed by Orgel and colleagues (Wu and Orgel 1992a), yet its mechanism had not previously been studied in detail. Interestingly, we found that there is no general acid-base catalysis in the rate-determining step of the reaction and the 3′-hydroxyl group is not deprotonated before attack; however, Mg^{2+} likely makes inner shell contacts during catalysis. From these findings we proposed a model wherein Mg^{2+} coordinates the 3′-hydroxyl of the primer and the phosphate of the incoming nucleotide to coordinate them for attack.

Chapter three continued to explore the mechanism of non-enzymatic RNA polymerization, but focused on the importance of the nucleotide-activating group. We confirmed and quantified a result originally found by Orgel and colleagues that the activating group of the nucleotide bound downstream to the reacting nucleotide affects the reaction rate (Wu and Orgel 1992b). Unexpectedly, we then found that free activating group can inhibit the polymerization reaction even though the reaction is irreversible. The inhibition did not appear to be competitive with nucleotide binding or due to binding to Mg^{2+}. Free activating group also inhibited hydrolysis of the activated nucleotide, although

less potently. These results have important implications for the *in situ* reactivation of nucleotides, which requires free activating group.

In chapter four we presented a novel synthesis for the 2′-amino-modified version of threose nucleic acid (2′-NH_2-TNA), a nucleic acid based on TNA, which many have suggested could be relavant to the origin of life (Schoning et al. 2000; Orgel 2000; Yu, Zhang, and Chaput 2012). Amino-modified nucleotides are attractive since they can polymerize more quickly than natural nucleotides due to the more nucleophilic amine (Schrum et al. 2009; Zhang et al. 2013) and since their lack of requirement for Mg^{2+} makes them compatible with fatty acid vesicles (Mansy et al. 2008). Although we expected the relatively rigid threose backbone to form a pre-organized template for efficient polymerization, we found that the amino-threose nucleotides polymerized more slowly than their amino-RNA counterparts and polymerized more slowly on a TNA template than on RNA or DNA. This finding suggests that a certain degree of flexibility is required for the complex to adopt a reactive conformation.

Finally, in chapter five we presented two new alkyne-modified initiator nucleotides that can be used to prime transcription by T7 RNA polymerase for subsequent coupling to azide-containing compounds like labels and affinity tags. One contained a terminal alkyne with a short $(CH_2)_2$ linker for copper(I)-catalyzed azide-alkyne cycloaddition coupling and the other contained a dibenzylcyclooctyne group with a PEG_4 linker for copper-free strain-promoted azide-alkyne cycloaddition coupling. Although these nucleotides are generally useful for efficiently modifying the 5′-terminus of transcripts, we developed them for an *in vitro* selection experiment designed to isolate a novel RNA-dependent RNA polymerase core. We proposed to couple an RNA library to a 2-methylimidazole-activated nucleotide

and then select for RNA sequences that could ligate that nucleotide to a primer. This selection could overcome three weaknesses in the R18 polymerase family of ribozymes: their lack of affinity for their primer-template substrate (Lawrence and Bartel 2003), the fact that their ligase core was originally selected with oligonucleotide substrates (Bartel and Szostak 1993), and their large size from having an accessory domain selected on top of a ligase core (Johnston et al. 2001).

Future Directions

The future issues and possible solutions for nucleic acid replication were discussed in chapter one and have been reviewed recently (Szostak 2012). It is not difficult to see why it is such a difficult problem. Nucleic acids are thermodynamically unstable to hydrolysis and the primer 3′-hydroxyl group is not a particularly good nucleophile. Indeed, it is no more reactive than water, which is at a concentration of 55 M. Therefore, the polymerization rate must be enhanced by a combination of activating the 3′-hydroxyl group (*e.g.*, general acid-base catalysis or deprotonation) and increasing its effective concentration by forming a primer-template-monomer complex. Our results in chapter two show that 3′-hydroxyl activation is likely minimal and the apparent K_D of the 2-MeImpG monomer is on the order of 10 mM. Despite these limitations the half-time for addition is reasonably quick at 24 min.

We have discussed a couple different approaches to enhancing the polymerization rate, including the use of amino-modified nucleotides and ribozyme catalysis. However, the mechanistic understanding gained through the experiments presented here can guide a third approach: the rational design of small molecule catalysts. Detailed studies of the

mechanism of RNA hydrolysis led to the development of a wide variety of small molecule catalysts for the reaction that use a combination of general acid-base catalysis and metal coordination (Morrow and Iranzo 2004; Lönnberg 2011; Cheng, Abhilash, and Breslow 2012). These catalysts have even been combined with sequence-specific DNA binding small molecules to give site-specific cleavage (Dervan 2001). Small molecule catalysts have been developed for many other biochemical reactions as well (Breslow 1995; Marchetti and Levine 2011).

In comparison, very few small molecule catalysts have been developed for non-enzymatic RNA polymerization. Intercalators, downstream binding oligonucleotides and polycations have all been used to enhance polymerization by strengthening binding (Renz, Lohrmann, and Orgel 1971; Horowitz et al. 2010; Deck, Jauker, and Richert 2011), and nucleophilic catalysts can enhance amino-nucleotide polymerization (Röthlingshöfer et al. 2008; Schrum et al. 2009; Zhang et al. 2013). Recent work from our group has shown that chelated Mg^{2+} can catalyze RNA polymerization, albeit less effectively than free Mg^{2+} (Adamala and Szostak). This finding raises the possibility of developing improved ligands analogous to those developed for RNA hydrolysis. For example, the net charge of the complex could be increased by attaching the Mg^{2+} to a polycation or by dimerizing the ligand to form a dinuclear complex. A particularly interesting starting point would be a dinuclear copper(II) complex that has previously be used to catalyze phosphodiester transesterification and the hydrolysis of 2′,3′-cyclic phosphates by coordinating to the phosphate and providing a bound hydroxide for general acid-base catalysis (Liu and Hamilton 1997; Liu, Luo, and Hamilton 1997). Ideally, the catalyst would be composed of a short peptide or another potentially prebiotic material to make it relevant to origins of life

research, and it should also allow for a sufficiently low concentration of divalent cation as to make the reaction compatible with fatty acid vesicles.

References

Adamala, K, and J W Szostak. "Non-Enzymatic Template-Directed RNA Synthesis Inside Model Protocells." *Science* in press.

Bartel, D P, and J W Szostak. 1993. "Isolation of New Ribozymes From a Large Pool of Random Sequences." *Science* 261 (5127) (September 10): 1411–1418.

Breslow, R. 1995. "Biomimetic Chemistry and Artificial Enzymes: Catalysis by Design." *Accounts of Chemical Research* 28 (June 4): 146–153.

Cheng, Liang, K G Abhilash, and Ronald Breslow. 2012. "Binding and Biomimetic Cleavage of the RNA Poly(U) by Synthetic Polyimidazoles." *Proceedings of the National Academy of Sciences USA* 109 (32) (August 7): 12884–12887. doi:10.1073/pnas.1210846109.

Deck, Christopher, Mario Jauker, and Clemens Richert. 2011. "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA." *Nature Chemistry* 3 (8) (August): 603–608. doi:10.1038/nchem.1086.

Dervan, PB. 2001. "Molecular Recognition of DNA by Small Molecules." *Bioorganic & Medicinal Chemistry* 9 (9): 2215–2235.

Horowitz, Eric D, Aaron E Engelhart, Michael C Chen, Kaycee A Quarles, Michael W Smith, David G Lynn, and Nicholas V Hud. 2010. "Intercalation as a Means to Suppress Cyclization and Promote Polymerization of Base-Pairing Oligonucleotides in a Prebiotic World." *Proceedings of the National Academy of Sciences USA* 107 (12) (March 23): 5288–5293. doi:10.1073/pnas.0914172107.

Johnston, WK, PJ Unrau, MS Lawrence, ME Glasner, and D P Bartel. 2001. "RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension." *Science* 292 (5520): 1319–1325.

Lawrence, MS, and D P Bartel. 2003. "Processivity of Ribozyme-Catalyzed RNA Polymerization." *Biochemistry* 42 (29): 8748–8755. doi:10.1021/bi0342281.

Liu, S H, and A D Hamilton. 1997. "Catalysis of Phosphodiester Transesterification by Dinuclear Cu(II) Complexes: the Role of the Second Cu(II) Ion." *Bioorganic & Medicinal Chemistry Letters* 7 (13): 1779–1784.

Liu, S, Z Luo, and A D Hamilton. 1997. "Rapid and Highly Selective Cleavage of Ribonucleoside 2',3'-Cyclic Monophosphates by Dinuclear Cu(II) Complexes."

Angewandte Chemie-International Edition in English 36 (23) (November 6): 2678–2680.

Lönnberg, Harri. 2011. "Cleavage of RNA Phosphodiester Bonds by Small Molecular Entities: a Mechanistic Insight." *Organic & Biomolecular Chemistry*: 1–17. doi:10.1039/c0ob00486c.

Mansy, Sheref S, Jason P Schrum, Mathangi Krishnamurthy, Sylvia Tobé, Douglas A Treco, and J W Szostak. 2008. "Template-Directed Synthesis of a Genetic Polymer in a Model Protocell." *Nature* 454 (7200) (July 3): 122–125. doi:10.1038/nature07018.

Marchetti, Louis, and Mindy Levine. 2011. "Biomimetic Catalysis." *ACS Catalysis* 1 (9) (September 2): 1090–1118. doi:10.1021/cs200171u.

Morrow, Janet R, and Olga Iranzo. 2004. "Synthetic Metallonucleases for RNA Cleavage." *Current Opinion in Chemical Biology* 8 (2) (April): 192–200. doi:10.1016/j.cbpa.2004.02.006.

Orgel, L E. 2000. "Origin of Life - a Simpler Nucleic Acid." *Science* 290 (5495): 1306–1307.

Renz, M, R Lohrmann, and L E Orgel. 1971. "Catalysts for the Polymerization of Adenosine Cyclic 2′,3′-Phosphate on a Poly(U) Template." *Biochimica Et Biophysica Acta* 240 (4): 463–471.

Röthlingshöfer, Manuel, Eric Kervio, Tanja Lommel, Ulrich Plutowski, Annette Hochgesand, and Clemens Richert. 2008. "Chemical Primer Extension in Seconds." *Angewandte Chemie-International Edition in English* 47 (32) (July 28): 6065–6068. doi:10.1002/anie.v47:32.

Schoning, KU, P Scholz, S Guntha, X Wu, R Krishnamurthy, and A Eschenmoser. 2000. "Chemical Etiology of Nucleic Acid Structure: the Alpha-Threofuranosyl-(3′->2′) Oligonucleotide System." *Science* 290 (5495): 1347–1351.

Schrum, Jason P, Alonso Ricardo, Mathangi Krishnamurthy, J C Blain, and J W Szostak. 2009. "Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2′-Amino-2′,3′-Dideoxyribonucleoside-5′-Phosphorimidazolide Monomers." *Journal of the American Chemical Society* 131 (40) (October 14): 14560–14570. doi:10.1021/ja906557v.

Szostak, J W. 2012. "The Eightfold Path to Non-Enzymatic RNA Replication." *Journal of Systems Chemistry* 3 (September 7): 2.

Wu, T F, and L E Orgel. 1992a. "Nonenzymatic Template-Directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides." *Journal of the American Chemical Society* 114 (1): 317–322.

Wu, T F, and L E Orgel. 1992b. "Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 2. Templates Containing Cytidine and Guanosine Residues." *Journal of the American Chemical Society* 114 (14): 5496–5501.

Yu, Hanyang, Su Zhang, and John C Chaput. 2012. “Darwinian Evolution of an Alternative Genetic System Provides Support for TNA as an RNA Progenitor.” *Nature Chemistry* 4 (3) (March): 183–187. doi:10.1038/nchem.1241.

Zhang, S, N Zhang, J C Blain, and J W Szostak. 2013. “Synthesis of N3′-P5′-Linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension.” *Journal of the American Chemical Society* 135 (January 7): 924–932. doi:10.1021/ja311164j.